The Thesis Committee for Silpakorn Dachanuwattana
Certifies that this is the approved version of the following thesis:

A History Matching Workflow Using Proxy-Based MCMC Applied in Tight Reservoir Simulation Studies

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

Kamy Sepehrnoori

Wei Yu

A History Matching Workflow Using Proxy-Based MCMC Applied in Tight Reservoir Simulation Studies

by

Silpakorn Dachanuwattana

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Engineering

The University of Texas at Austin

August 2017

Dedication

To my parents, **Vinit** and **Busara**, and my brothers, **Jiraphat** and **Akaradej**

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my graduate supervisor, Dr. Kamy Sepehrnoori, for the continuous guidance and encouragement during these past two years. The guidance received in the weekly meeting has greatly shaped my research direction. I also truly appreciate the S.P. Yates Memorial Endowment for Graduate Fellowships he nominated me even before I had asked. After all, it has been a rewarding experience to work under his supervision.

My sincere appreciation goes to Dr. Wei Yu for his dedication, suggestion and discussion. I gain a tremendous knowledge from his expertise in unconventional reservoirs, hydraulic fracturing, and reservoir simulation. I am really thankful for his valuable time to review my thesis and teach me an academic writing skill.

I would also like to thank to Marut Wantawin, Pável Zuloaga Molero, and Yifei Xu for their collaboration and insightful ideas for my researches. In addition, I would like to extend my thanks to all my officemates for knowledge sharing and memorable experiences. It has been my privilege to be a part of this wonderful team.

Finally, I am extremely grateful to my family in Thailand and U.S.A for their endless supports in all of my pursuits including my lifetime goal of obtaining this graduate degree. My life is truly blessed to have you all.

Abstract

A History Matching Workflow Using Proxy-Based MCMC Applied in Tight Reservoir Simulation Studies

Silpakorn Dachanuwattana, M.S.E.

The University of Texas at Austin, 2017

Supervisor: Kamy Sepehrnoori

Reservoir simulation for tight reservoirs often requires characterization of hydraulic and natural fracture networks in the reservoir model. Microseismic data reveals approximate boundary of the fracture networks but its direct application of stimulated rock volume (SRV) fails short to capture fracture connectivity and fracture conductivity, which significantly dominate well performance. Embedding discrete fractures in reservoir simulation is thus required to attain more realistic reservoir behavior. However, using local grid refinement (LGR) to model discrete fractures is computationally expensive. Even more challenging is generating multiple realizations of the fracture-embedded reservoir models during history-matching. Not only one simulation but extensive series of simulations are required to deal with complex geometry of fractures as well as other uncertain parameters. However, recent developments in a methodology called Embedded Discrete Fracture Model (EDFM) have overcome the computational complexity using discrete fractures in reservoir simulations.

In this thesis, we develop an efficient assisted history matching (AHM) workflow using proxy-based Markov chain Monte Carlo (MCMC) algorithm and integrate the workflow with the EDFM preprocessor. To improve the efficiency, the optimal proxy is studied by comparing the performance of four types of proxies: quadratic polynomial, cubic polynomial, k-nearest neighboring (KNN), and kriging under various contexts such as different measurement errors. The results show that kriging proxy is more accurate than KNN proxy and cubic proxy. The quadratic proxy was the least accurate in our evaluations. However, if larger measurement error is introduced, the distinction between accuracy of the four proxies becomes less clear in spite of their different computational costs.

Incorporating these findings, the proxy-based MCMC workflow is developed and implemented in conjunction with the EDFM to history match a shale oil well in Vaca Muerta formation to demonstrate the application of the workflow. The microseismic data are accounted to constrain the uncertain geometries of the fractures. The integrated workflow can successfully and efficiently history match the actual shale-oil well with complex fractures. Not only the uncertainties of reservoir properties are narrowed down but the posterior likelihood of fracture geometry scenario is also attained after history matching.

We also compare the proxy-based MCMC workflow with the direct MCMC and a commercial history matching software in terms of accuracy and efficiency. It is found that the direct MCMC cannot find enough solutions to construct the posterior probability density (PPD) in an efficient manner. For the commercial software, it can find solutions faster than the proxy-based MCMC. However, the former is stuck in the local minima, thus resulting in an invalid PPD. Ultimately, the proxy-based MCMC workflow provides the most accurate history matching results with efficient manner for this tight oil reservoir.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 MOTIVATION

Uncertainty quantification is crucial for decision making in reservoir developments and the process is very challenging. Reservoir simulation, which is a common tool to quantify the production forecasting uncertainty, inherits several uncertainties of geological model (e.g., structure and stratigraphy), reservoir properties (e.g., porosity and permeability), and fluid properties. History matching is therefore required to narrow down the uncertainty ranges – by choosing only the reservoir realization that is conditioned the measured data. Since the problem is generally under-determined, not only one but several realizations satisfy history matching. In addition, the sampling frequencies of these solutions must follow the posterior probability density (PPD) constrained by historical data. Since a history matching problem is highly nonlinear and deals with multi-dimensional parameters in general, constructing the PPD is extremely computationally expensive and practically impossible to achieve by manually history matching. The assisted history matching (AHM) therefore becomes the mainstream approach in petroleum industry.

There have been numerous studies of AHM algorithms since at least last two decades in the industry. Some authors used stochastic optimizers such as pilot-point method (Barker et al. 2001), differential evolution (Hajizadeh et al. 2010), genetic algorithms (Zhang et al. 2013), particle swarm optimization (Vazquez et al. 2015), and distributed Gauss-Newton method (Gao et al. 2016). However, the posterior PDFs constructed by stochastic optimizers are biased to the optimization algorithms themselves (Goodwin 2015), and not sampled from the true PPD. Other two prominent methods are ensemble Kalman filter (EnKF) method (Evensen et al. 2007; Li et al. 2009) and randomized maximum likelihood (RML) method (Chen et al. 2016). These two methods

are capable of constructing the true posterior PDF for linear problems. Unfortunately, reservoir problems are highly non-linear; hence, EnKF and RML are only considered as approximating methods for history matching. Rejection method and direct MCMC are known of ability to construct the true posterior PDF (Oliver et al. 2008). Each of them involves a criterion for accepting or rejecting reservoir-model realizations generated from a stochastic process. The main drawback of direct MCMC is its expensive computational cost. Rejection method requires even more cost especially when dealing with high quantity of uncertain parameters. It is important to note that each AHM algorithm has its own trade-off in terms of efficiency and accuracy. Researchers have made significant progress to overcome the drawbacks of AHM algorithms using various methods. The techniques include the application of Design of Experiment (DOE) to reduce the number of parameter (Schaaf et al. 2009), the hybrid between two methods such as EnKF and MCMC (Emerick et al. 2010), and the implementation of a proxy model (Slotte et al. 2008; Goodwin 2015).

While most studies of the AHM applications have been conducted on conventional reservoirs (Barker et al. 2001; Yeten et al. 2005; Ma et al. 2006; Slotte et al. 2008; Busby et al. 2009; Mohamed et al. 2009; Emerick et al. 2010; Elsakout et al. 2015; Olalotiti-Lawal et al. 2015), relatively few studies were performed on unconventional reservoirs (Yang et al. 2015; Wantawin et al. 2017). Compared to conventional resources, history matching unconventional resources may face some extra challenges such as highly uncertain fracture patterns, complex geomechanics properties, and nanoscale adsorption. Data acquisition is another challenging issue especially when numerous wells are drilled during the early development stage. The significant difference between the two types of reservoirs could make the application of an AHM workflow on an unconventional reservoirs not equally appropriate as that on conventional reservoirs. Therefore, AHM workflows should be tested on real unconventional reservoirs to check the robustness and efficiency.

Moreover, reservoir simulation model for shale plays often requires characterization of hydraulic and natural fracture networks. The fracture modelling often requires microseismic data but the direct application of stimulated rock volume (SRV) fails short to capture fracture connectivity and fracture conductivity, which significantly dominate well performance. Embedding discrete fractures in reservoir simulation is thus necessary to provide more realistic reservoir behavior but the task is computationally challenging. Local grid refinement (LGR) or unstructured gridding have been used to add discrete fractures into a reservoir model but the methods are intrusive to the model and anticipate with high computational cost. However, recent developments in a methodology called Embedded Discrete Fracture Model (EDFM) have overcome the computational complexity using discrete fractures in reservoir simulations. Therefore, developing a history matching framework that combines the strength of EDFM will provide an efficient tool to reduce uncertainty of tight reservoirs with complex fracture geometry

1.2 Research Objectives

The aim of this study is to develop an efficient AHM workflow that is suitable for tight reservoirs with complex fractures. First, we study and understand the trade-offs of different types proxies used in proxy-based MCMC. In addition, this study addresses the impacts of initial points, measurement errors, and termination criteria on history matching process. The recommended workflow is applied to the synthetic and real field cases of a tight reservoir to demonstrate the implementation. Also, the performance of the workflow is also compared with a commercial history matching software. Finally, combined with EDFM, the workflow is highly effective to history match the real field case with complex fracture geometry.

1.3 THESIS OUTLINE

The thesis consists of six chapters and is organized as follows. Chapter 2 reviews previous works covering various AHM algorithms and their trade-offs. It also introduces different proxy types used to accelerate AHM algorithms, and the challenges to history match tight reservoirs.

Chapter 3 presents the investigation of the proxy-based MCMC workflow. The study is performed on a synthetic simple case to compare the accuracy of the four proxy types under different measurement errors. In addition, the impact of initial points is studied and the termination criterion is recommended.

In Chapter 4, the proposed workflow using proxy-based MCMC is demonstrated on a real field case of a shale oil well in Vaca Muerta formation. Moreover, we conduct the comparative study between the AHM workflow, the direct MCMC, and a commercial history matching software in terms of efficiency and accuracy.

In Chapter 5, history matching of the same shale-oil well is performed. Microseismic data is used to provide more realistic fracture network. The AHM workflow is implemented in conjunction with EDFM preprocessor to history match the shale oil well with uncertain and complex fracture geometry.

Finally, Chapter 6 summarizes the key findings and provides recommendations for the future work that potentially improves the performance of the proposed proxy-based MCMC workflow to history match tight reservoirs with complex fractures.

Chapter 2: Literature Review

2.1 Assisted History Matching (AHM) Techniques

AHM algorithms have been studied for more than two decades in oil and gas industry. These algorithms are difficult to clearly categorize because novel methods have been continuously invented and some methods are hybridization of others. However, AHM approaches might be broadly classified into three categories: (1) Optimization-based methods, (2) Bayesian methods, and (3) Ensemble Kalman Filter (ENKF).

2.1.1 Optimization-Based Methods

Optimization-based methods use optimization algorithms that search for local or global minima. Numerous optimization algorithms have been implemented in AHM, including steepest descent (Chen et al. 1974), pilot-point method (Barker et al. 2001), differential evolution (Hajizadeh et al. 2010), genetic algorithm (Cheng et al. 2008), particle swarm optimization (Vazquez et al. 2015), and distributed Gauss-Newton method (Gao et al. 2016). An AHM workflow does not involve only the algorithm. The parameterization process, screening process, and structure of a workflow also contribute to the performance, hence worthwhile studying. Some optimization-based workflows are exemplified as follows.

Landa et al. (2003) used an optimization-based algorithm combined with kriging proxy for AHM in the field case of a conventional oil reservoir. The objective function is the single misfit value based on the three-phased flow rates and shut-in pressures. 9 tuning parameters were defined including pore volume multipliers and transmissibility multipliers. The workflow is an iterative process where four specific points were added in each iteration. The four points include (1) minimum-error point, (2) maximum-kriging-variance point, (3) a local point near the minimum-error point, and (4) a local point near

the maximum-kriging-variance point. The search for these four points is performed by optimization-based algorithms including genetic algorithm and Gauss-Newton algorithm. The authors run the workflow for 18 iterations for the total of 72 simulation points before using kriging to build the final proxy. Several points were then randomly drawn from this final proxy. Only the points of which the proxy objective functions are lower than the prescribed target were regarded as the history matching solutions.

Cheng et al. (2008) used genetic algorithms (GA) to history match the Tengiz carbonate oil field, Kazakhstan. The objective function was the misfits between the pressure survey data and the corresponding simulated responses. The history matching was conducted stage-by-stage; at global level, then flow unit level, followed by well level. In each stage, parameterization and sensitivity analysis were performed to select the heavy-hitter parameters. The number of heavy-hitter parameters in each stage varied from 11 to 40 parameters and GA, in conjunction with proxy, was used to find the history matching solutions. The solutions in a stage were used directly in the next stage but they were used narrow down the uncertainty ranges. Totally, the workflow required 2300 simulation runs to find about 400 solutions. Cluster analysis revealed that 370 out of these 400 realizations were considered redundant so only the unique 30 realizations were accepted for probabilistic forecasting. On average, 77 simulations were required to find one unique solution. An interesting feature of this workflow is the structure approach. Dividing a history matching task into several subtasks, that are influenced by lower-dimensional parameters, results in the higher efficiency.

Gao et al. (2016) used the distributed gauss-newton (DGN) method for history matching in a synthetic field case varying 235 input parameters. The DGN is an iterative process able to find multiple local minima and is modified from the Gaussian Newton method. The modifications include starting with multiple initial points and using local

quadratic model to gain information from nearby points. The workflow was run for 50 iterations for the total of about 28000 simulation runs. 900 history-matching solutions were found or 31 simulation runs were required to find one history-matching solution.

Optimization-based methods tend to find history-matching solutions with relatively lower computational cost. However, optimization-based algorithms are not designed for uncertainty quantification. The solutions found by the algorithms can be biased even though some post-processing schemes such as clustering or weighting are applied. Uncertainty quantification based on a set of wrongly distributed solutions presumably prone to bias, thus leading to wrong decisions. To overcome this pitfall, Bayesian methods are introduced due to their abilities to sample the solutions from the correct distribution.

2.1.2 Bayesian Methods

Bayesian methods apply Bayes Theorem to sample history-matching solutions. A distinct attribute of these methods is the use of prior knowledge about uncertain parameters. The Bayesian solution is not only a single solution but the entire posterior probability density (PPD) accounting for prior knowledge, data error, and model error. Bayesian methods that have been used as AHM include Markov chain Monte Carlo (MCMC), acceptance-rejection method (AR) (Yang et al. 2015) and randomized maximum likelihood (RML) (Gao et al. 2005). While MCMC and AR are capable of sampling the correct PPD in a non-linear problem such as history matching, RML can only approximate the PPD since the method is designed for a linear problem.

Several classes of MCMC have been applied as AHM such as Metropolis-Hasting (MH) (Slotte et al. 2008), Gibb's Sampling (Hamdi et al. 2017), and Hamiltonian MCMC (Goodwin 2017). MH is probably the class of MCMC used most frequently in the literature

and simply referred as MCMC by many researchers. In this thesis, MH will be interchangeably referred as MCMC, unless otherwise specified.

Liu et al. (2003) used a synthetic problem investigating the performance of the direct MCMC as an AHM algorithm. The flow scenario was 1D, single-phased flow in heterogeneous reservoir. The case has the analytical solution which simulate the flow much faster than numerical simulation. The goal was to history match BHP where the tuning parameters are permeability and porosity. Despite successful history matching, the authors reported 100 million simulations were required for the MCMC chain to converge.

Slotte et al. (2008) used a proxy-based MCMC algorithm for AHM on a conventional, relatively homogeneous oil and gas reservoir in the Heidrun Field, Norwegian Sea. The global objective function was the misfit between the 331 measurements of production history and the corresponding simulated responses. 56 tuning parameters were selected as the input for history matching. Kriging proxy was built for each measurement in the objective function. The stop criteria was when the uncertainty of the objective function is of the same order as the spread in the objective function values for the sampled ensemble of the final proxy. The workflow was run iteratively for a total of 300 simulations. The correlation factor between the estimated value and simulated values for the final proxy was 0.52 before the final proxy was used to generate probabilistic forecasts. Despite the correlation factor of only 0.52, the accuracy of the proxy would be improved if longer iterations had been run. The PPD was then constructed by using MCMC to explore the final proxy.

2.1.3 EnKF

EnKF is an iterative process to estimate the state of a dynamical system from a series of noisy measurements. The algorithm is derived from the simple Kalman filter

(Kalman 1960) and has been widely applied in various reservoir engineering applications including AHM (Aanonsen et al. 2009). EnKF assumes Gaussian probability distributions and linear relationship between input and response parameter. Applying the filter in history matching problems where the relationships are highly non-linear and the distributions are non-Gaussian would yield suboptimal or even invalid results. To address the limitations, numerous impressive researches have been introduced such as the works of Gao et al. (2005), Emerick et al. (2011), and Tavakoli et al. (2014).

Each AHM has its own advantages and disadvantages in terms of accuracy versus computational cost. Hybridization between methods is another promising strategy to synergize the advantages of methods in the aforementioned three categories. For instance, the integration of EnKF and MCMC was introduced by Emerick et al. (2010). However, none of AHM method is manifested as the standard practice in reservoir engineering community. This is probably because the selection of an AHM method to use also depends on tolerance of the accuracy target, nonlinearity of the problem, and availability of computational capacity, which is problem-specific. Hence, the method of choice is not always the most sophisticated approach but probably the one that is fit for purpose.

2.2 PROXY

A proxy is a mathematical model to approximate the response of an unmeasured point in multi-dimensional parameter space given measured points. In history-matching context, proxies have often been used to substitute simulation runs that are much more computationally expensive. Interchangeably, a proxy is also called a surrogate model, a response surface model, or a multi-dimensional interpolation. It can be classified as an exact or an inexact proxy (Johnston et al. 2002b). An exact proxy predicts the responses exactly identical to the observed data at any measured points. In contrast, an inexact proxy

does not necessarily predict the similar responses to the observed data. Examples of the exact proxies are kriging, thin-plate spline, radial basis function, and KNN while examples of the inexact ones are polynomial regression and artificial neural network.

Four types of proxies comprising of quadratic polynomial, cubic polynomial, KNN, and kriging are studied in this thesis. Kriging is an exact proxy used in AHM by several authors such as Landa et al. (2003) and Goodwin (2015). Although kriging has been reported to be highly accurate in many literatures, its computational cost is expensive and proportional to the cube of the number of measured points. This high cost makes kriging impractical to apply for an overwhelmingly large set of data. On the other end of the spectrum, polynomials are computationally cheaper and have also been used in AHM by several authors such as Zhang et al. (2013) and Wantawin et al. (2016). However, being an inexact proxy, polynomial proxy compromises the accuracy of proxy in some situations. KNN is an exact proxy like kriging but requires much less computational cost. Despite being rarely used in the petroleum industry, it can be a substitute for kriging if computational cost is a limiting factor. A brief explanation for each proxy type is provided below.

2.2.1 Quadratic Polynomial

Quadratic polynomial uses the regression method to identify the relationship between independent variables and a dependent response. Although quadratic proxy is inexact, the polynomial has been widely used to construct proxies either in history matching or optimization application due to its simplicity and cheap computational cost. The equation for quadratic polynomial is shown below.

$$\hat{z}(\vec{\theta}_0) = \beta_0 + \sum_{i=1}^{n_d} \beta_i \theta_i + \sum_{i=1}^{n_d} \sum_{j>i}^{n_d} \beta_{ij} \theta_i \theta_j + \sum_{i=1}^{n_d} \beta_{ii} \theta_i^2 \,, \tag{2.1}$$

where $\vec{\theta}_0 = [\theta_1, \theta_2, \theta_3, \ldots, \theta_{n_d}]$ is a vector of independent variables; $\hat{z}(\vec{\theta}_0)$ is the predicted response; $\beta_0, \beta_i, \beta_{ij}$, and β_{ii}, are the coefficients calculated based on the least square method.

2.2.2 Cubic Polynomial

Similar to quadratic polynomial, cubic polynomial uses the regression method to fit the relationship between independent variables and a dependent response. Cubic polynomial however has more regression coefficients than quadratic polynomial thus can better fit a more complex relationship. The equation for cubic polynomial is shown below

$$\hat{z}(\vec{\theta}_0) = \beta_0 + \sum_{i=1}^{n_d} \beta_i \theta_i + \sum_{i=1}^{n_d} \sum_{\substack{j=1 \\ j>i}}^{n_d} \beta_{ij} \theta_i \theta_j + \sum_{i=1}^{n_d} \beta_{ii} \theta_i^2 + \sum_{i=1}^{n_d} \sum_{j>i}^{n_d} \sum_{k>j}^{n_d} \beta_{ijk} \theta_i \theta_j \theta_k + \sum_{i=1}^{n_d} \sum_{j>i}^{n_d} \beta_{iij} \theta_i^2 \theta_j + \sum_{i=1}^{n_d} \sum_{j>i}^{n_d} \beta_{ijj} \theta_i \theta_j^2 + \sum_{i=1}^{n_d} \beta_{iii} \theta_i^3, \tag{2.2}$$

where $\vec{\theta}_0 = [\theta_1, \theta_2, \theta_3, \ldots, \theta_{n_d}]$ is a vector of independent variables; $\hat{z}(\vec{\theta}_0)$ is the predicted response; $\beta_0, \beta_i, \beta_{ij}, \beta_{ii}, \beta_{ijk}, \beta_{iij}, \beta_{ijj},$ and β_{iii}, are the coefficients calculated based on the least square method.

2.2.2 K-Nearest Neighboring (KNN).

KNN is a technique for pattern classification (Cover et al. 1967). To predict the response of an unmeasured point, the KNN searches the nearest k measured points then averages their responses. The choice of averaging methods is arbitrary but should provide a higher weight for a nearer measured point. One of common-averaging methods is the inverse distance weighted averaging method described Eq. 2.3. The KNN using this averaging method is also called inverse distance weighted (IDW) interpolation.

$$\hat{z}(\vec{\theta}_0) = \sum_{i=1}^{k} \lambda_{i0} z(\vec{\theta}_i), \tag{2.3}$$

where $\vec{\theta}_0$ is a vector of independent variables, $\hat{z}(\vec{\theta}_0)$ is the predicted response, $\vec{\theta}$ is one of the nearest k measured points, $z(\vec{\theta}_i)$ is the corresponding measured response, and λ_{i0} is the weight function between $\vec{\theta}_0$ and $\vec{\theta}_i$ calculated by the following expression:

$$\lambda_{i0} = \frac{\left|\vec{\theta}_i - \vec{\theta}_0\right|^{-1}}{\sum_{i=1}^{k} \left|\vec{\theta}_i - \vec{\theta}_0\right|^{-1}}, \tag{2.4}$$

2.2.4 Kriging

Kriging is an interpolation technique for multi-dimensional parameter space originally used in geostatistics. Similarly to the KNN, kriging forms weight from surrounding measured points to predict the response of an unmeasured point and kriging is also an exact proxy. However, the weight function of kriging is more complex than the KNN. Not only based on the distance, but the weight function is also based on the spatial structure of the data (Johnston et al. 2002a), which is called the variogram. A classical method to construct a variogram is proposed by Matheron (1962):

$$\hat{\gamma}(\vec{h}) = \frac{1}{2\left|N(\vec{h})\right|} \sum_{N(\vec{h})} \left(z(\vec{\theta}_i) - z(\vec{\theta}_j)\right)^2, \tag{2.5}$$

where $\hat{\gamma}(\vec{h})$ is the variogram, $N(\vec{h}) = \{(i,j): \vec{\theta}_i - \vec{\theta}_j = \vec{h}\}$, and $\left|N(\vec{h})\right|$ is the number of distinct pairs of $N(\vec{h})$.

Constructing the variogram is a critical step before predicting a response using kriging. There are various methods to construct a variogram other than Eq. 2.5, but we do not cover in this study. Conceptually, a variogram is using available measured data to construct a function that quantifies the dissimilarity of response ($\hat{\gamma}$) as a function only of

the increment location vector $\vec{\theta}_i - \vec{\theta}_j$. Kriging also uses Eq. 2.3 to predict the response. In this study we use ordinary kriging which the weight function, λ_{i0}, is calculated by Eq. 2.6 (Isaaks et al. 1989).

$$\begin{pmatrix} \lambda_{10} \\ \lambda_{20} \\ \vdots \\ \lambda_{k0} \\ \mu_0 \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1k} & 1 \\ c_{21} & c_{22} & \cdots & c_{2k} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ c_{k1} & c_{k2} & \dots & c_{kk} & 1 \\ 1 & 1 & \dots & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} c_{10} \\ c_{20} \\ \vdots \\ c_{k0} \\ 1 \end{pmatrix}, \tag{2.6}$$

where $c_{ij} = \lim_{\alpha \to \infty} \hat{\gamma}\left(\alpha\left(\vec{\theta}_i - \vec{\theta}_j\right)\right) - \hat{\gamma}\left(\vec{\theta}_i - \vec{\theta}_j\right)$, μ_0 is the Lagrange multiplier.

2.3 CHALLENGES OF HISTORY MATCHING FOR TIGHT RESERVOIRS

History matching for tight reservoirs presents additional challenges compared to conventional reservoirs. The distinct complexity of tight reservoir includes highly uncertain fracture patterns, complex geomechanics properties, and complex fluid transport mechanisms in nanopores. Data availability is another challenging issue, especially when numerous wells are drilled during the early development stage but the data acquisition is limited.

Hydraulic fracturing and horizontal drilling are the key technologies enabling economic production of tight reservoirs for the last decade. Unlike the simple planar fracture in the classic theory, hydraulic fracture patterns in shale reservoirs can be much more complex due to the interactions with pre-existing natural fractures (Cipolla et al. 2010). The fracture connectivity can be vastly different depending on the different geometries. Figure 2.1 (Warpinski et al., 2009) displays possible fracture patterns in tight reservoirs, where the red dots represent the well location and the blue lines represent the fracture geometries. As illustrated, the fracture geometries may be roughly categorized as

the complex fracture, the complex fracture with fissure opening, and the complex fracture network.

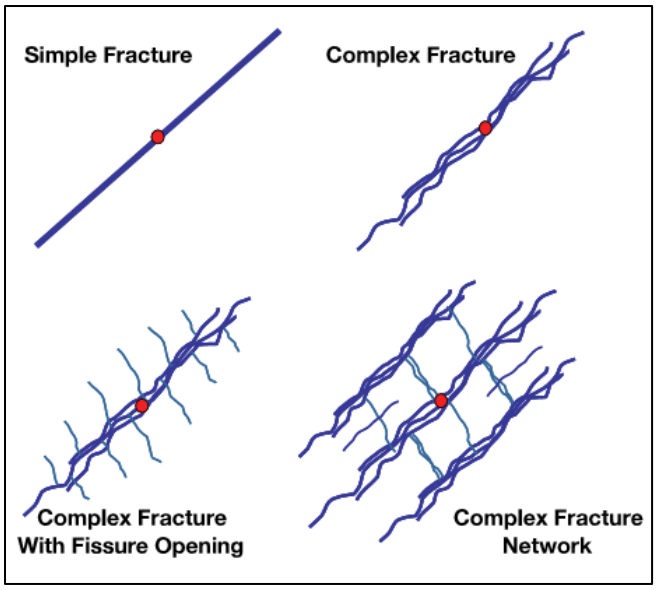


Figure 2.1: Schematic of fracture pattern scenarios

Microseismic monitoring is the most widely used technique that provides the most information about the fracture geometry (Warpinski et al. 2013). Several approaches have been developed in oil and gas industry to interpret the fracture geometry from microseismic data. The microseismic event hypothetically originates by the two mechanisms: (1) the induced stress near the fracture tip, and (2) the leak-off of the high-pressured fracturing fluids into natural fractures or other weakness planes (Warpinski et al. 2013). The microseismic events occur very close to the fracture propagation during the fracturing stimulation, the microseismic event pattern therefore could be used to determine the fracture geometry. One approach to interpret microseismic data is using the total reservoir volume encompassed by the cloud of microseismic events to assess the effectiveness of the fracturing stimulation. The total volume commonly called the stimulated rock volume (SRV) is displayed in Fig. 2 as for an example. The hypothesis is that the higher SRV is

correlated to the better production performance. While the SRV is easy to calculate and its hypothesis sounds intuitive, the SRV does not adequately describe the production performance (Maxwell et al. 2009). This is because the SRV does not account for the important fracture properties such as connectivity, conductivity, and geometry which are found to be crucial for the well performance and the reservoir behavior. This raises the importance of other approaches that are capable of generating the fracture pattern discretely while honoring microseismic data and the other constraints such as geomechanics.

One approach to discretely populate fractures is using fracture propagation model (Cipolla et al. 2012; Taleghani et al. 2013). The fracture geometry can be generated with the integration between the model and the microseismic data. The reliability of the resultant fracture pattern essentially depends on the quality of the microseismic data as well as other key parameters for the fracture propagation. However, microsesimic data generally carry some uncertainty in its acquisition and interpretation processes. Also, there is usually limited information of the key parameters that govern fracture propagation such as in-situ stress, modulus, and natural-fracture pattern (Warpinski et al. 2013). As a consequence, characterizing and history matching tight reservoirs are extremely challenging.

Different techniques have been used to history match unconventional reservoirs. Samandarli et al. (2011) used the analytical solution and semi-analytical solution to history match fracture permeability, matrix permeability, and fracture half-length in Barnett shale gas wells. Clarkson et al. (2014) demonstrated the application of an analytical solution to history match multi-fractured horizontal well in liquid-rich resources. Analytical or semi-analytical solutions are quick AHM approaches for some simple reservoir cases. Nevertheless, the assumptions of these methods are overly-simplified, thus prohibiting their applications in more complex reservoirs.

More sophisticated schemes have also been implemented for AHM in unconventional reservoirs. Yin et al. (2011) used the proxy-based GA to history match a synthetic shale gas well. In addition to the production data, SRV was used in their history matching as an additional constraint to reduce the uncertainty of the tuning parameters. Nejadi et al. (2015) used EnKF to history match a shale gas well in the Horn River resource play, Canada. Their workflow applied discrete-fracture-network (DFN) model before upscaling to dual porosity model for reservoir simulation. However, as mentioned in section 2.1, the distribution of history matching from GA and EnKF is prone to bias and potentially lead to inaccurate uncertainty quantification. Hence, the thesis will revolve around the study of the proxy-based MCMC as AHM since the MCMC algorithm provides correct PPD and its efficiency can be improved with the implementation of a proxy.

Chapter 3: Investigation of the AHM Workflow using Proxy-Based MCMC in Synthetic Cases

This chapter describes two case studies using synthetic models: one is synthetic simple case and another is synthetic field case. The first case is used to visualize the evolution of the response surfaces over MCMC iterations. Also, the case is used to evaluate the impacts of various contexts, such as measurement error, on the performance of the AHM workflow. The second case is used as the demonstration of the AHM workflow on a more complex problem.

3.1 Synthetic Simple Case

We use this case as the basis to address following four questions: (1) What is the accuracy of each proxy? (2) When should we terminate iteration? (3) What is the accuracy of each proxy if the observed data carry measurement errors? (4) Does the number of simulation points at initial iteration affect the AHM workflow efficiency? The case description is a hydraulically-fractured vertical well symmetrically placed in a rectangular homogeneous reservoir, as illustrated in Figure 3.1. The oil rate profile is given and the goal is to match the historical bottomhole pressure (BHP) profile by varying the two uncertain parameters: fracture half-length (x_f), and permeability (k).

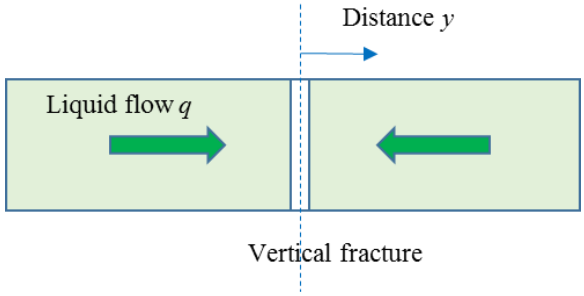


Figure 3.1: Side view of the rectangular reservoir with a hydraulically-fractured vertical well

This problem has an analytical solution available as follows (Behmanesh et al. 2014):

$$p_D(y_D,t_D)=\pi[2\sqrt{\frac{t_D}{\pi}}\cdot\exp(-\frac{y_D^2}{4t_D})-y_D\cdot erfc(\frac{y_D}{2\sqrt{t_D}})], \tag{3.1}$$

$$p_D=\frac{2\pi kh(p_i-p)}{qB\mu}, t_D=\frac{kt}{\phi\mu C_t x_f^2}, y_D=\frac{y}{x_f}, \tag{3.2}$$

where p is the reservoir pressure at time t and at perpendicular distance y to the fracture, p_i is initial reservoir pressure, k is permeability, q is surface oil rate, B is formation volume factor, μ is oil viscosity, ϕ is porosity, C_t is total compressibility, and x_f is fracture half-length. The basic reservoir and fracture parameters in this case study are listed in Table 3.1, unless otherwise specified.

Table 3.1: Summary of reservoir and fracture parameters for the synthetic simple case

Parameter	Symbol	Unit	Value
The perpendicular distance	y	ft	0.01
Initial reservoir pressure	p_i	psi	8000
Permeability	k	md	0.05
Formation volume factor	B	bbl/STB	1.2
Viscosity	μ	cp	5
Porosity	ϕ	-	0.2
Total Compressibility	C_t	1/psi	1.0×10-6
Fracture half-length	x_f	ft	250

We arbitrarily chose reference parameters: $x_{f,ref}$ and k_{ref} to synthesize the measured BHP profile using Eqs. 3.1 and 3.2, as shown in Figure 3.2. It is assumed that there are no measurement errors for oil flow rates, hence the measured rate is equal to the physically actual rate. The history-matching goal is to find all the realizations that minimize BHP root-mean-square-error (RMSE). For any realizations (x_f and k), we use Eqs. 4.1 and 4.2

to generate the simulated BHP profiles then use Eq. 4.3 to calculate BHP RMSE. We could even generate the true BHP RMSE response surface of every realization of x_f and k as shown in Figure 3.3. It is important to note that this true BHP RMSE response is never used for constructing the proxy but only used as the benchmark to validate the accuracy of proxy.

$$\text{BHP RMSE}(x_f, k) = \sqrt{\left\{\sum_{j=1}^{n}\left[\text{BHP}(x_f, k, t_j) - \text{BHP}(x_{f,ref}, k_{ref}, t_j)\right]^2\right\} / n}, \tag{3.3}$$

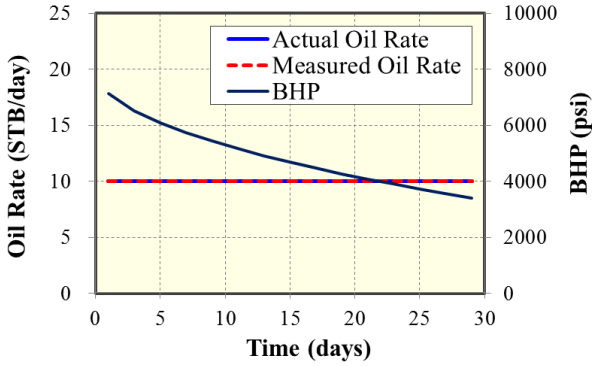
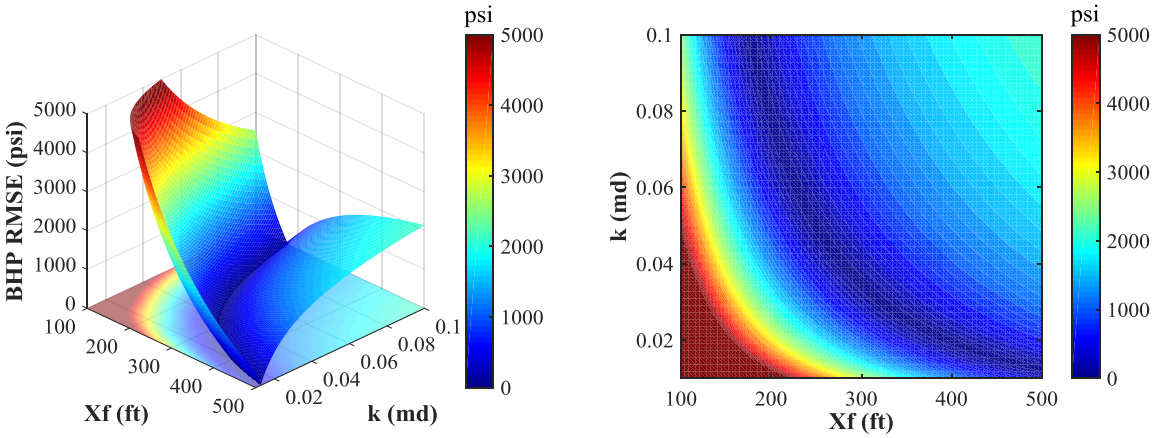


Figure 3.2: BHP profiles, actual oil rate, measured oil rate without measurement errors



(a) Surface plot of the true BHP RMSE response surface

(b) Contour plot of the true BHP RMSE response surface

Figure 3.3: The true BHP RMSE response surface from brute force method

3.1.1 What Is The Accuracy of Each Proxy?

The methodology to compare accuracy of each proxy is summarized as the flowchart shown in Figure 3.4. First, we use Latin Hypercube to randomly sample 25 initial points from the domains of x_f and k . The BHP profiles of these points are then generated using Eqs. 3.1 and 3.2. After that, Eq. 3.3 is used to calculate the BHP RMSE of each point. Now that we have the BHP RMSE responses of every initial points, the proxies are used to estimate the RMSE surface for all unmeasured points in the entire domain. The proxies based on the 25 sampled points may not adequately represent the true response surface. We therefore draw additional 25 points using MCMC and iterate the process to improve the quality of proxy. In this study, we terminate the workflow when completing 20 iterations.

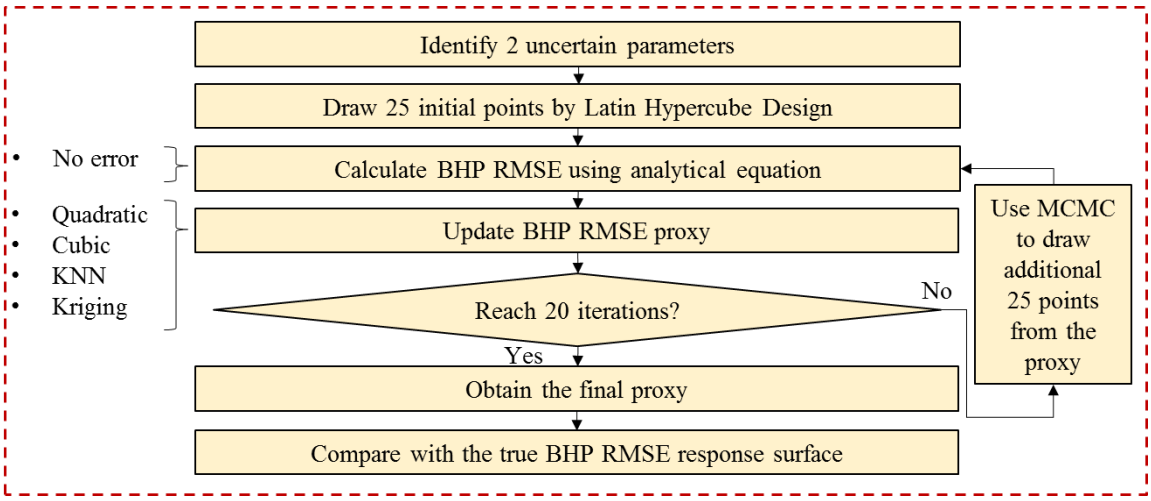


Figure 3.4: Workflow of the history matching study to compare the performance of the four different proxy types in the synthetic simple case

The resulting BHP RMSE response surfaces of the 1st, 2nd, and 20th iterations are shown as the contour maps in Figure 3.5, Figure 3.6, and Figure 3.7 respectively. The red dots in these figures represent the simulation points cumulatively drawn by MCMC to construct the proxy response surfaces. As illustrated, the quadratic and the cubic proxies

evolve only subtly whereas the KNN and the kriging proxies evolve continuously over iterations. We compare the proxy response surfaces at the 20th iteration, in Figure 3.7, with the true response surface in Figure 3.3(b). As can be seen, the quadratic proxy does not appear to resemble the true response surface but the cubic proxy does better. The KNN and kriging proxies represent the true response surface most accurately especially in the low-BHP-RMSE area in which the history-matching solutions are located.

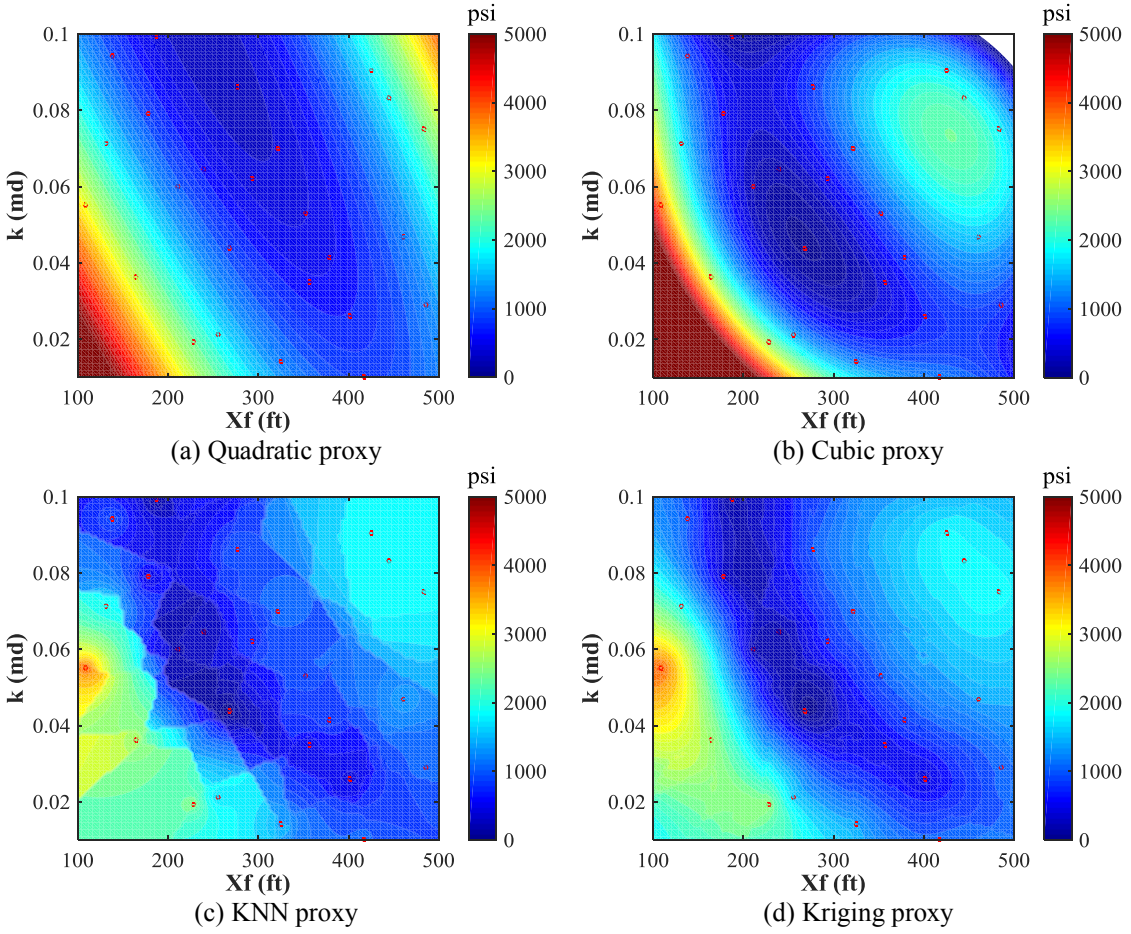
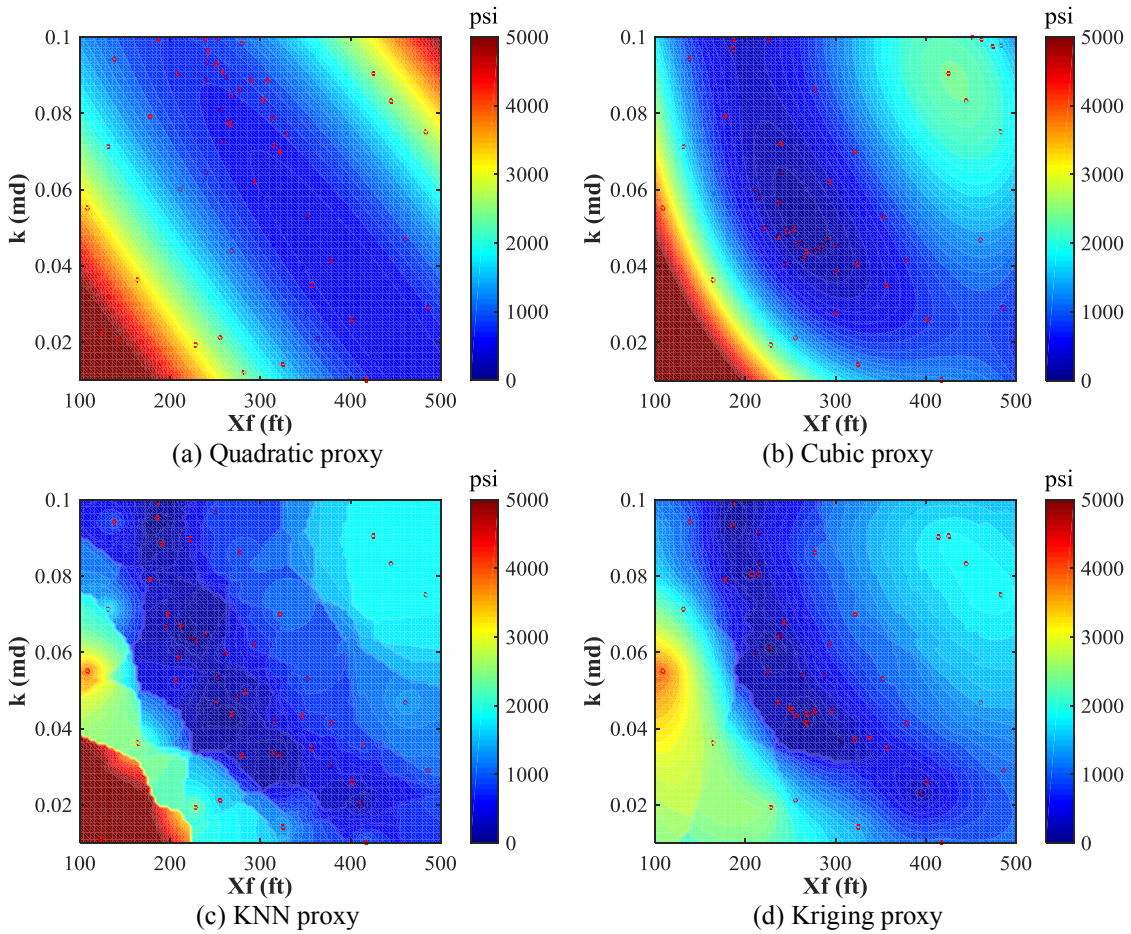


Figure 3.5: Contour plots of the proxy BHP RMSE response surfaces and the corresponding cumulative simulation points at 1st iteration



(a) Quadratic proxy (b) Cubic proxy

(c) KNN proxy (d) Kriging proxy

Figure 3.6: Contour plots of the proxy BHP RMSE response surfaces and the corresponding cumulative simulation points at 2nd iteration

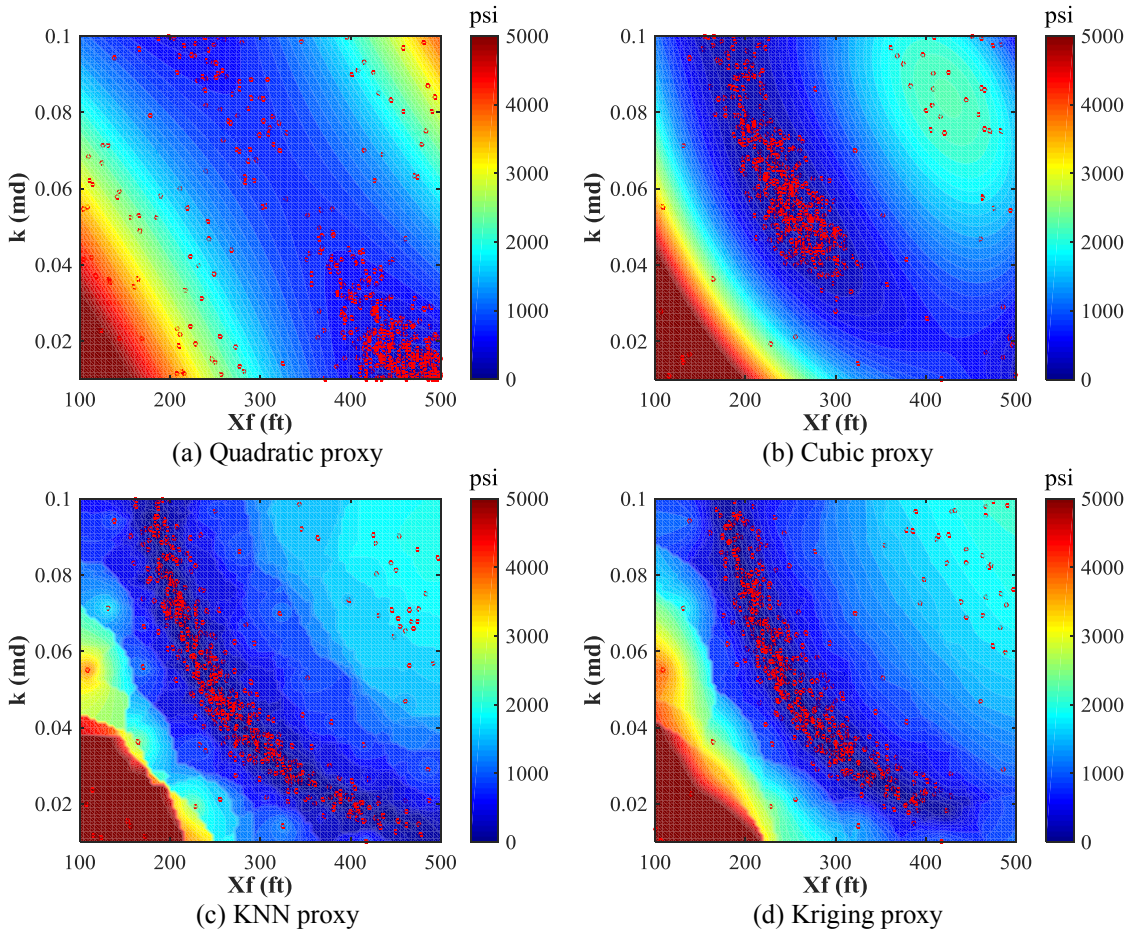


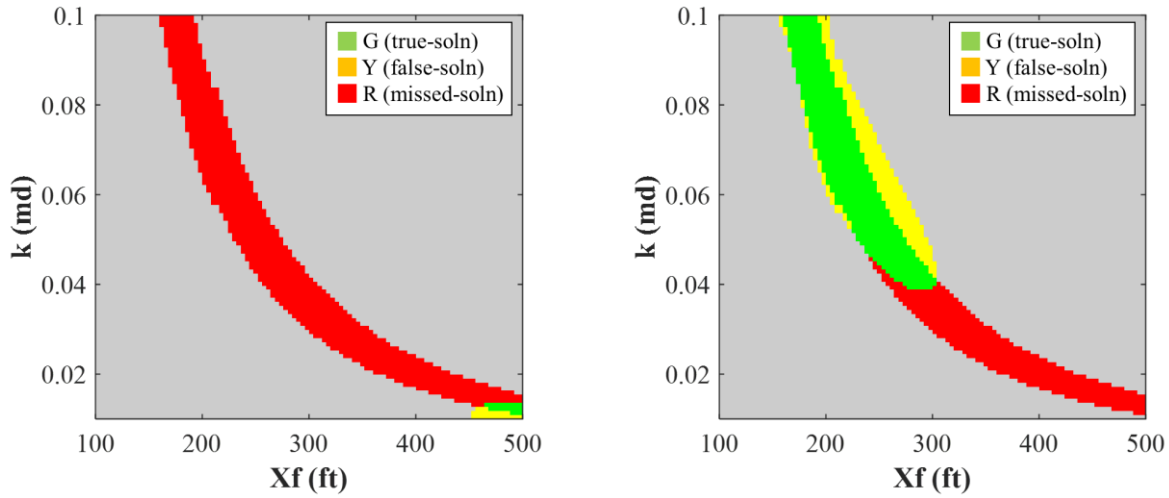
Figure 3.7: Contour plots of the proxy BHP RMSE response surfaces and the corresponding cumulative simulation points at 20th iteration

The prescribed BHP RMSE tolerance is set to be 300 psi. Any point that has proxy BHP RMSE lower than this target value is considered a history-matching solution. We compare the solutions domain from the proxy response surface against that from the true response surfaces. Inspired by Landa et al. (2003) the metric to measure the accuracy of proxy (AOP) is quantified by Eq. 3.4. We compare the AOP of the four proxy response surfaces in Figure 3.7 against the true response surface in Figure 3.3 and illustrate the comparison in Figure 3.8. Quantitatively, the kriging proxy is the most accurate, with 79.4% accuracy. The KNN proxy is the second one with 70.0% accuracy. The third most

accurate proxy is cubic proxy with 45.1% accuracy. The quadratic proxy is least accurate one with 1.7% accuracy.

$$\text{AOP} = \frac{G}{G+Y+R}, \tag{3.4}$$

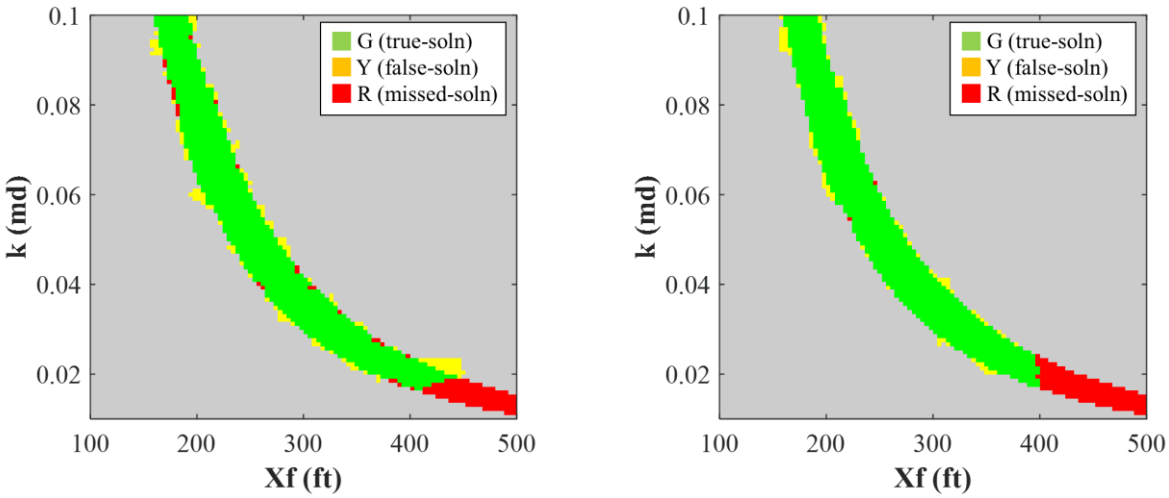
where G is the true-solution domain, where both the proxy and the true response surfaces are mutually lower than the target; Y is the false-solution domain meaning, where the proxy response surface is lower than the target but the true one is not; and R is the missed-solution domain, where the true response surface is lower than the target but the proxy one is not.



(a) Quadratic proxy

(b) Cubic proxy

Figure 3.8: continued next page.



(c) KNN proxy (d) Kriging proxy

Figure 3.8: The accuracy of proxy maps at 20th iteration (soln in the plots represent solution)

There are two issues about Eq. 3.4 noteworthy to mention. Firstly, it compares only the area of the parameter domains but does not account for the sampling density which varies over the area. We however use this metric due to its simplicity and the assumption that the sampling density is not significantly different over a narrow area of the parameter domains. Secondly, when we got the history-matching solutions proposed by the proxy response surface for a real field case. It is easy to disregard any false solutions ("Y" domain) simply by re-checking against running reservoir simulations. On the other hand, it is far more difficult to realize the occurrence that the proxy response surface misses some true solutions ("R" domain). The issue may be alleviated by introducing an indicator that indirectly measures the quality of proxy, which will be discussed in later part of this chapter.

We use Eq. 3.4 to calculate and plot the AOP with respect to iteration, shown in Figure 3.9. As can be seen, the kriging proxy has the highest AOP compared to the other proxies for almost all iterations. The KNN proxy gradually increases AOP from 2.6% in

the first iteration up to 70.0% which ends up the second most accurate proxy in the 20th iteration. Another notable result is that the AOP curves of kriging and KNN proxies are concave indicating the improvement rates of AOP diminish over iterations. Reservoir engineers may have to weigh the trade-off between incremental accuracy and computational cost to decide whether or not to continue for the next iteration. Although the cubic proxy has the highest AOP among all the proxies in the first iteration, this proxy has the slower improvement rate of AOP than kriging and KNN. The cubic proxy is eventually the third most accurate the proxy at the last iteration. For the quadratic proxy, we can see that its AOP is almost zero for all iterations suggesting that the quadratic proxy is inadequate to describe the response surface for this history-matching problem. However, we will further investigate the accuracy of quadratic proxy as well as the other proxies when introducing measurement errors later.

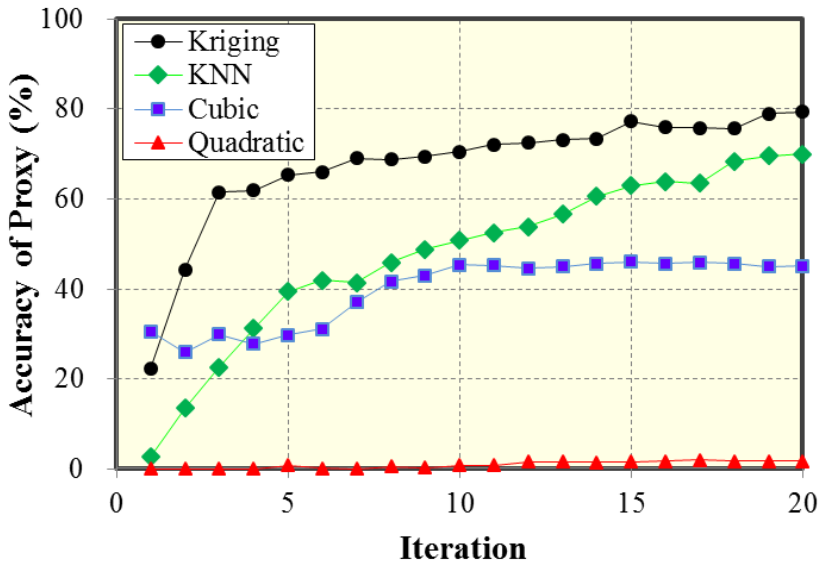


Figure 3.9: Accuracy of proxy using the different proxy types vs iteration

3.1.2 When Should We Terminate Iteration?

It is reasonable to terminate iteration in an AHM workflow once the AOP meets the prescribed accuracy target. However the AOP itself cannot be explicitly calculated unless all the solutions, such as Figure 3.3, are known beforehand. In practice, we perhaps calculate some metrics that indirectly measure how accurately a response surface in the current iteration can predict the response of the next iteration. Unlike the AOP calculation in Eq. 3.4 that compares every point in the domain, these metrics compare only some points sampled by MCMC. One of such the metrics is R^2 that is defined below:

$$R_j^2 = 1 - \frac{\sum_{i=1}^{n} (\mathrm{RMSE}_{proxy,j,i} - \mathrm{RMSE}_{actual,j+1,i})^2}{\sum_{i=1}^{n} (\mathrm{RMSE}_{actual,j+1,i} - \overline{\mathrm{RMSE}}_{actual,j+1})^2}, \tag{3.5}$$

$$\overline{\mathrm{RMSE}}_{actual,j+1} = \frac{1}{n} \sum_{i=1}^{n} \mathrm{RMSE}_{actual,j+1,i}, \tag{3.6}$$

where n is the number of simulation points in an iteration, $RMSE_{proxy}$ is the proxy RMSE and $RMSE_{actual}$ is the actual RMSE, the subscript j and $j+1$ depict the iteration numbers, the subscript i depicts a simulation point out of the total n points in an iteration.

For instance, we plot the actual RMSE vs proxy RMSE in the 1st, 10th, and 19th iterations of the synthetic simple case using KNN proxy as shown in Figure 3.10. The R^2's of the three iterations are calculated using Eqs. 3.5 and 3.6. We can observe that the proxy in the later iterations more accurately predict the actual RMSE. This qualitative observation is consistent with the quantitative metric R^2 that shows increasing values: 0.04, 0.75 and 0.98 for the 1st, 10th, and 19th iteration, respectively.

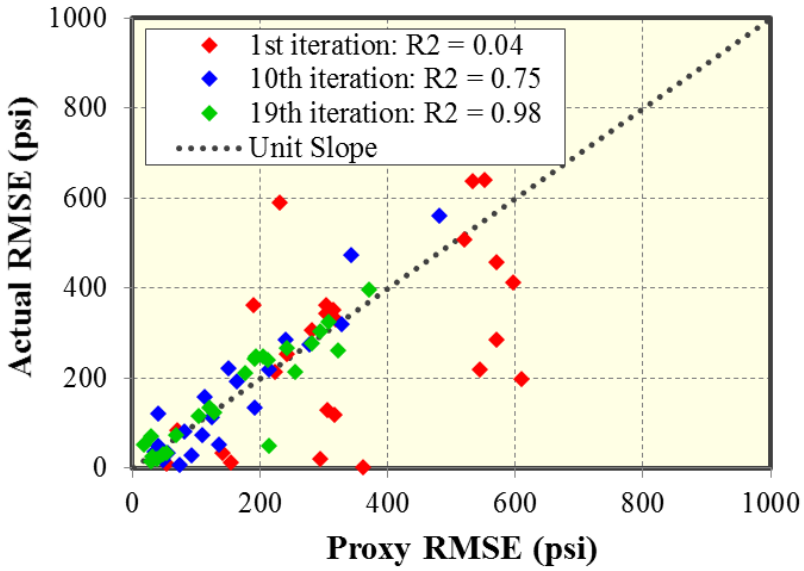
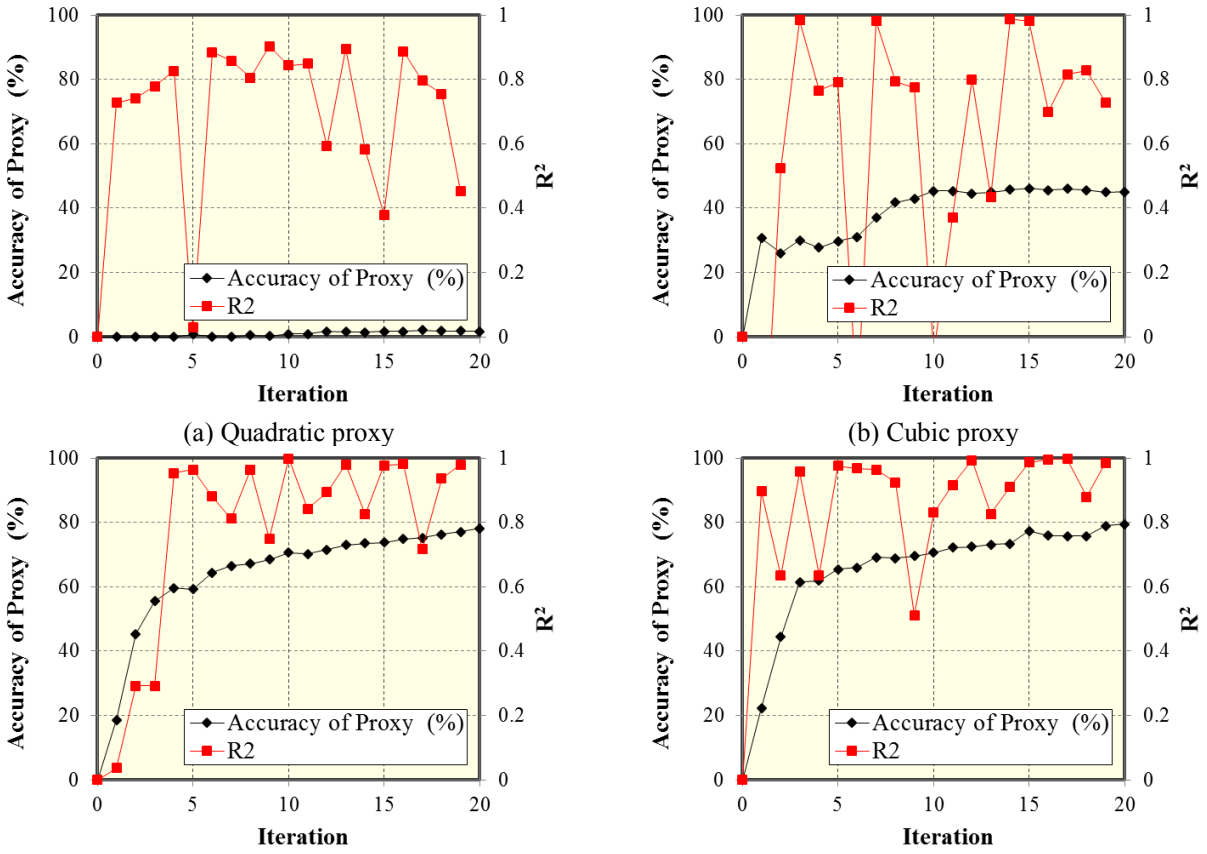


Figure 3.10: Actual RMSE vs Proxy RMSE using KNN proxy

We extend this application by calculating the R^2 for every iteration of each of the proxies, then overlay the R^2-vs-iteration with the AOP-vs-iteration as illustrated in Figure 3.11. It can be seen from Figure 3.11 (c) and (d) that the AOP trends of KNN and kriging proxies are essentially correlated with the trends. However, the AOP trends of the quadratic and cubic proxies are not really correlated with the trends as displayed in Figure 3.11(a) and (b). The low correlations are partly due to the limited number of regression coefficients in both of the quadratic and cubic proxies. With a certain number of the coefficients, adding new simulation points after having run several iterations will not really change the proxy response surface because the coefficients are still dominated by the influence of previous simulation points. With this reason, the R^2 of quadratic and cubic proxies barely indicate the accuracy of proxy. As can be seen in Figure 3.11(a) that accuracy of quadratic proxy is very low despite the high of 0.8. In contrast, KNN and kriging proxies are data-exact approximations that literally utilize every single simulation points to update their response

surfaces. Their AOP will consequently improve continuously when using more simulation points to update the proxies.



(a) Quadratic proxy

(b) Cubic proxy

(c) KNN proxy

(d) Kriging proxy

Figure 3.11: AOP and R^2 of the four different proxy types vs iteration

With the demonstration of R^2 as the guideline for the AHM workflow termination shown, a few remarks are recommended as follows:

- Using the R^2 as the guideline is appropriate for exact proxies such as kriging and KNN proxies, but it is not recommended for inexact proxies such as quadratic or cubic proxies.

- The number of simulation points in an iteration (n) should be statistically high enough. This is the reason why we use 25 points for our study.
- Although the AOP is well correlated to R^2, the two metrics are different things; R^2 of 1.00 does not necessarily mean the AOP of 100%.
- The R^2 vs iteration plot will not always be monotonic. It is in fact expected to have some fluctuation due to the nature of stochastic sampling. We would therefore monitor its trend rather than point-by-point.

3.1.3 Does The Initial Iteration Impact On The AHM Workflow Efficiency?

In the previous workflows, we use Latin Hypercube to sample initial 25 points. The number of initial points is chosen arbitrarily in general and it seems that no guideline is established. Hence, we investigate if the number of the initial points impacts on the AHM workflow efficiency. The problem description is similar to that in the section 4.2.1 but the workflow is slightly different and summarized in Figure 3.12. We prepare six cases with the different numbers of initial simulation points: 5, 10, 25, 50, 100, and 200 points, and evaluate the AOP more frequently at every five simulation points until the workflow reaches the cumulative number of 500 points. The proxy used is KNN as opposed to kriging because the computational time of kriging is too high in our experience.

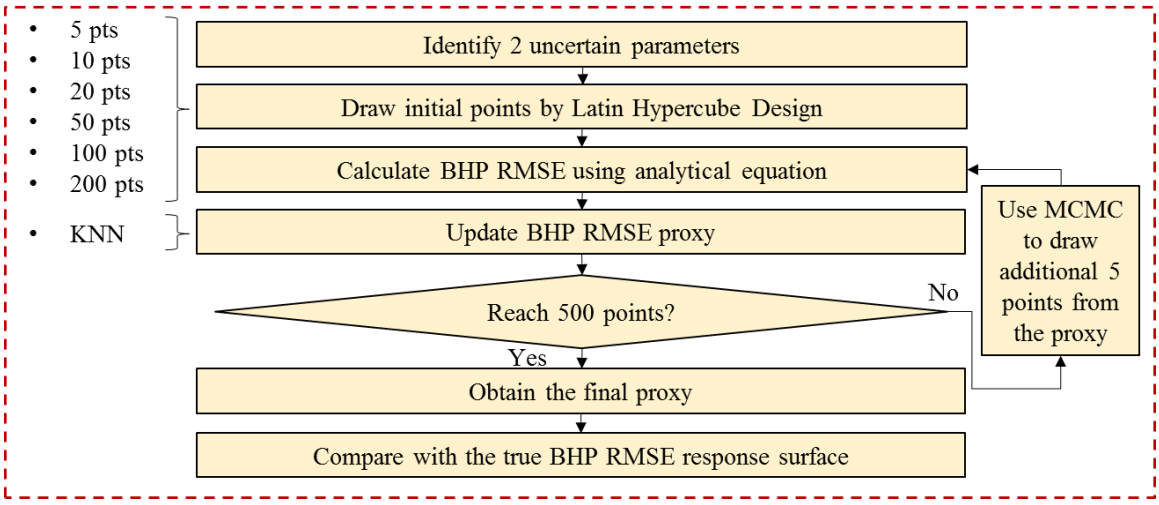


Figure 3.12: Workflow of the history matching study to compare the performances of the different initial points in the synthetic simple case

The prescribed BHP RMSE tolerance is set to be 100 psi. We use the stricter tolerance than 300 psi, previously used, so that the history-matching solutions are more difficult to be found by the AHM workflow. This difficulty will allow us to more obviously notice the impact of the initial simulation points on the workflow efficiency. We use Eq. 3.4 to calculate and plot the AOP vs iteration for all the six cases. As shown in Figure 3.13, at cumulative 200 simulation points, the initial-200-point case has the least AOP compared the other cases. Conversely, the initial-5-point case, does not outperform all the other cases. This suggests that too low or too high number of initial points result in the suboptimal AHM workflow efficiency. The difference of AOP between each case is substantial; considering the accuracy target of 60% for an example, the initial-5-point case achieves the target at 340 simulation runs while the initial-200-point case can achieve at 470 simulation runs. The latter case spends 38% more simulation runs, so the optimal number of initial simulation points is worth further investigation.

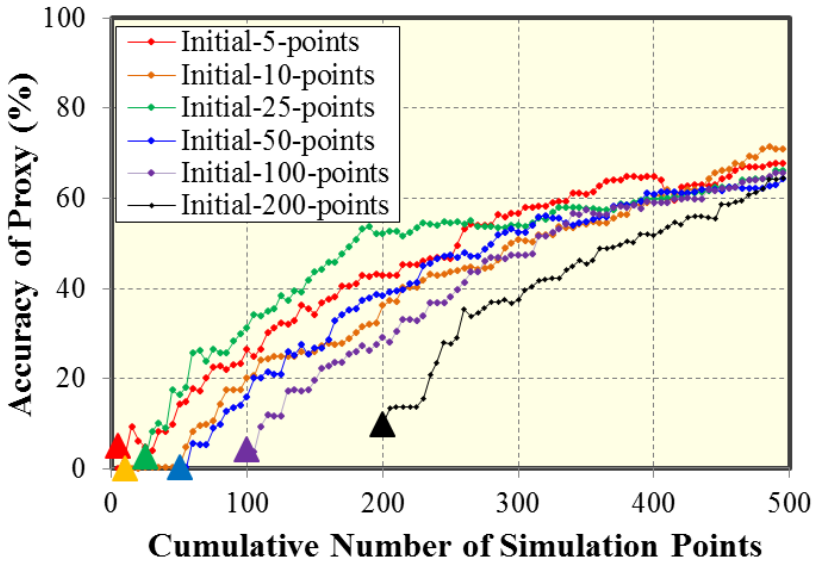


Figure 3.13: Accuracy of proxy using the different initial points vs iteration

3.1.4 What is the accuracy of each proxy if the observed data carry measurement errors?

In section 3.1.1, the measured oil rate is identical to the actual oil rate since we assume no measurement errors. Nevertheless, measurement errors are known to be embedded in the real data and we investigate their effect on the accuracy of each proxy. To this end, we synthesize two subcases: a small-error case and a big-error case. The small-error case has the actual oil rate with a small deviation from the measured oil rate. The deviation is randomly synthesized using Gaussian distribution with 1 STB/day standard deviation. The big-error case has the actual rate with a big deviation from the measured rate, randomly synthesized using Gaussian distribution with 5 STB/day standard deviation. This latter case intentionally exaggerates the magnitude of the error in order to make its impact more obvious. Eqs. 3.1 and 3.2 are still the bases to generate BHP profiles for these two cases but the temporal superposition principle is also required to account for the effect of changing rates. The production histories for the small-error case and the big-error case

are illustrated in Figure 3.14 and Figure 3.15, respectively. Reservoir engineers are supposed to see only the measured oil rate profiles, but not the actual oil rate that is physically related to BHP profiles. They must use the measured oil rate profiles, which carry the measurement errors, as the input to history match the BHP profiles.

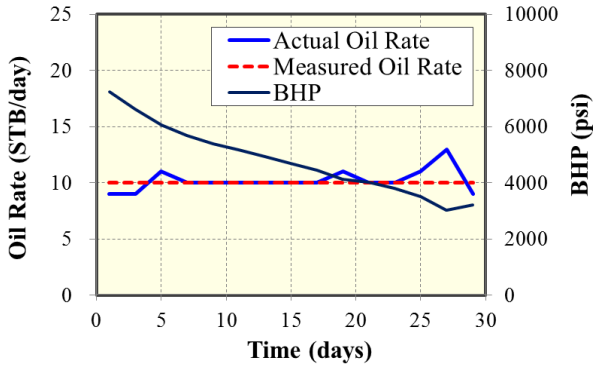


Figure 3.14: BHP profiles, actual oil rate, and measured oil rate of the small-error case

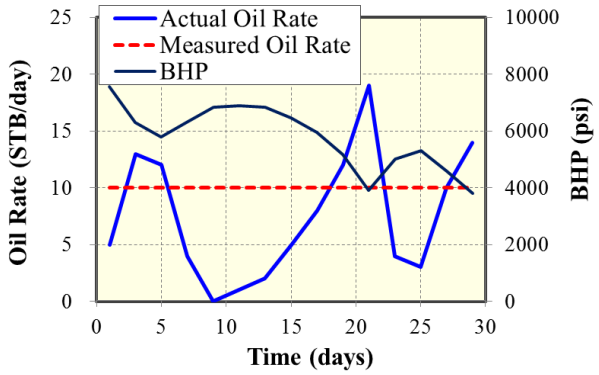
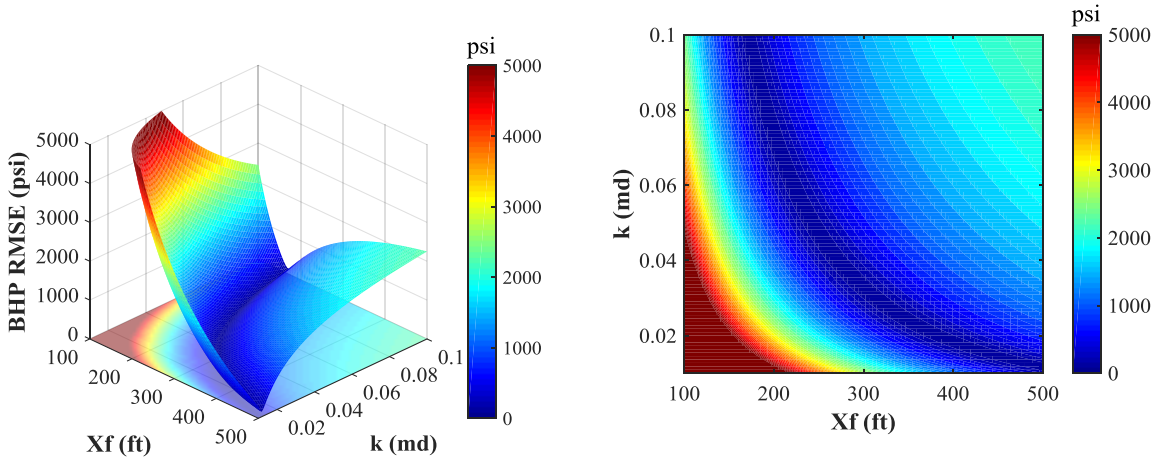


Figure 3.15: BHP profiles, actual oil rate, and measured oil rate of the big-error case

For the benchmarks to compare with the proxies in the subsequent step, we use Eq. 3.3 to calculate the true BHP RMSE response surface of the small-error case, as displayed in Figure 3.16. That of the big-error case is displayed in Figure 3.17. The true response surface of the small-error case in Figure 3.16 is trivially different from that of the no-error

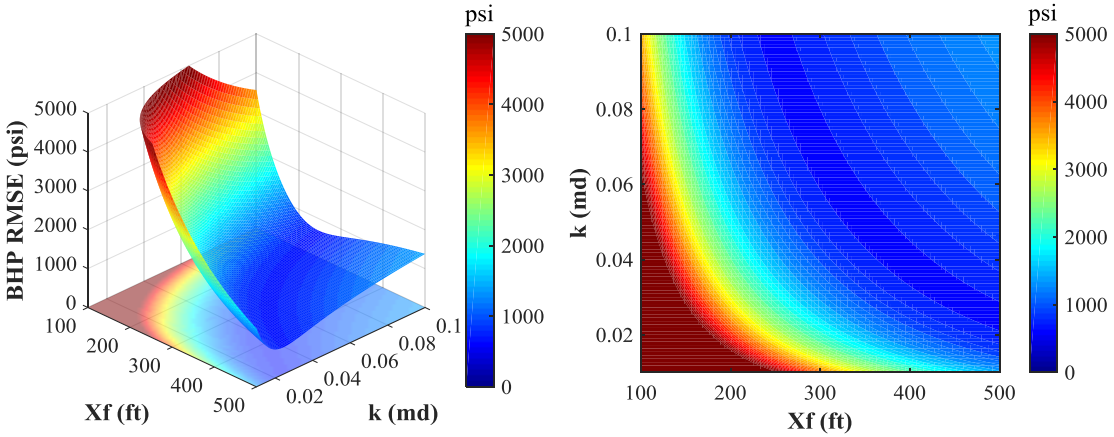
case in Figure 3.3 but the response of the big-error case in Figure 3.17 is substantially different. Moreover, the response surface of the big-error case does not have any BHP RMSE lower than 500 psi since the low responses are smeared by the impact of the measurement error. In fact, the lowest BHP RMSE can be found in this big-error case is BHP RMSE of 690 psi. We use the same prescribed RMSE tolerance of 300 psi, so the RMSE target is 990 psi in the big-error case.



(a) Surface plot of the true BHP RMSE response surface

(b) Contour plot of the true BHP RMSE response surface

Figure 3.16: The true BHP RMSE response surface of the small-error case from brute force method



(a) Surface plot of the true BHP RMSE response surface

(b) Contour plot of the true BHP RMSE response surface

Figure 3.17: The true BHP RMSE response surface of the big-error case from brute force method

We apply the AHM workflow in Figure 3.18 for the big-error case. The resulting response surfaces of the 1st, 2nd, and 20th iteration are shown in Figure 3.19, Figure 3.20, and Figure 3.21, respectively. The comparison between the proxy response surfaces in Figure 3.21 and the true response surface in Figure 3.3 are displayed in Figure 3.22. We use Eq. 3.4 to calculate and plot the AOP vs iteration in Figure 3.23.

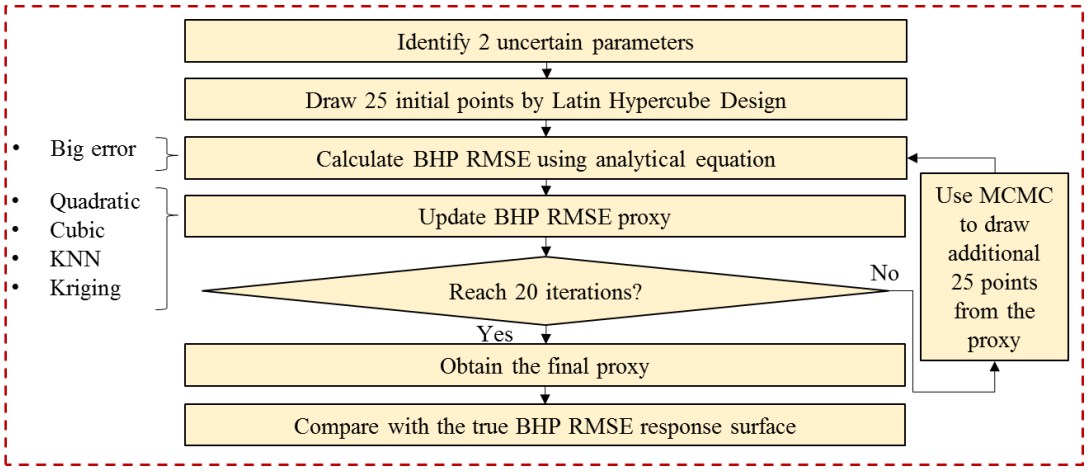
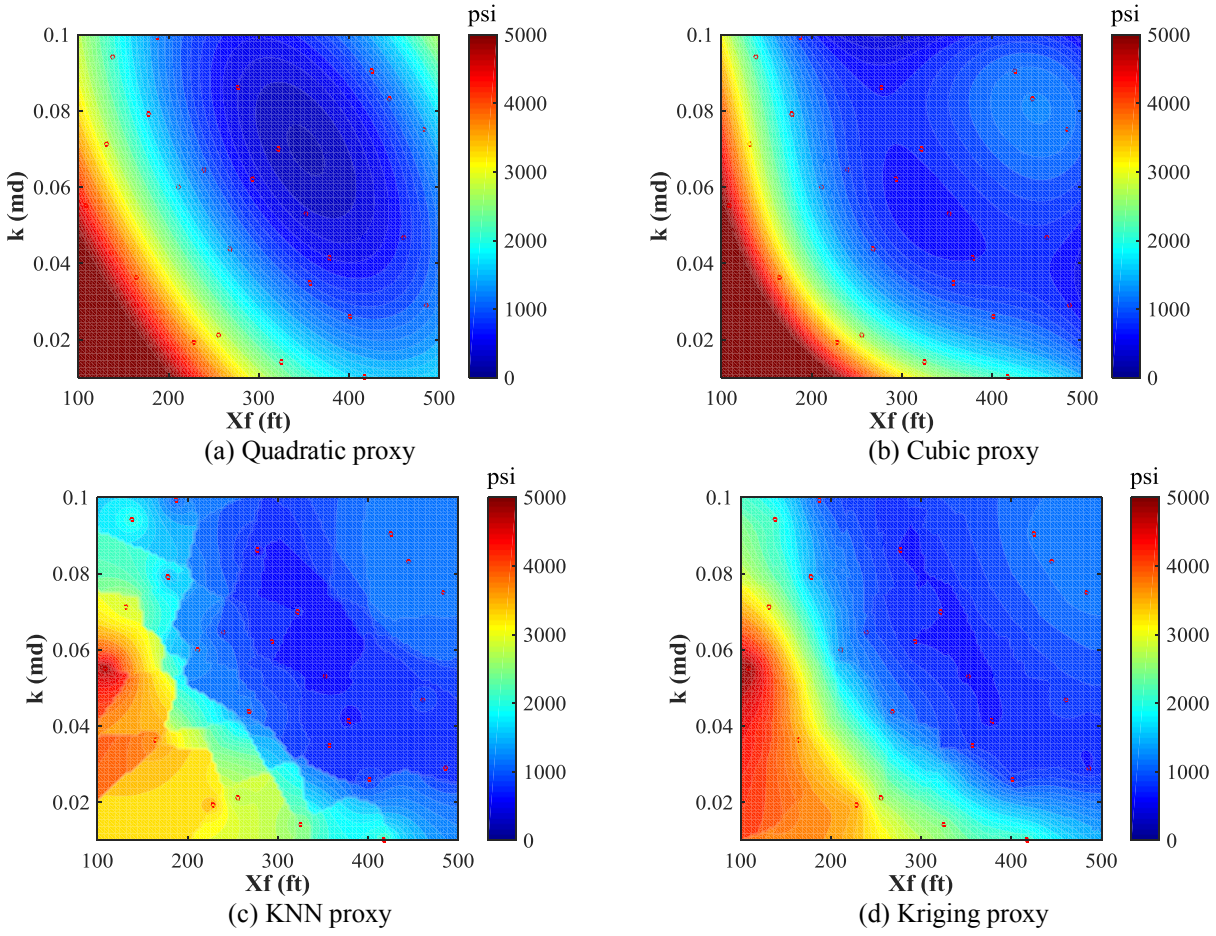


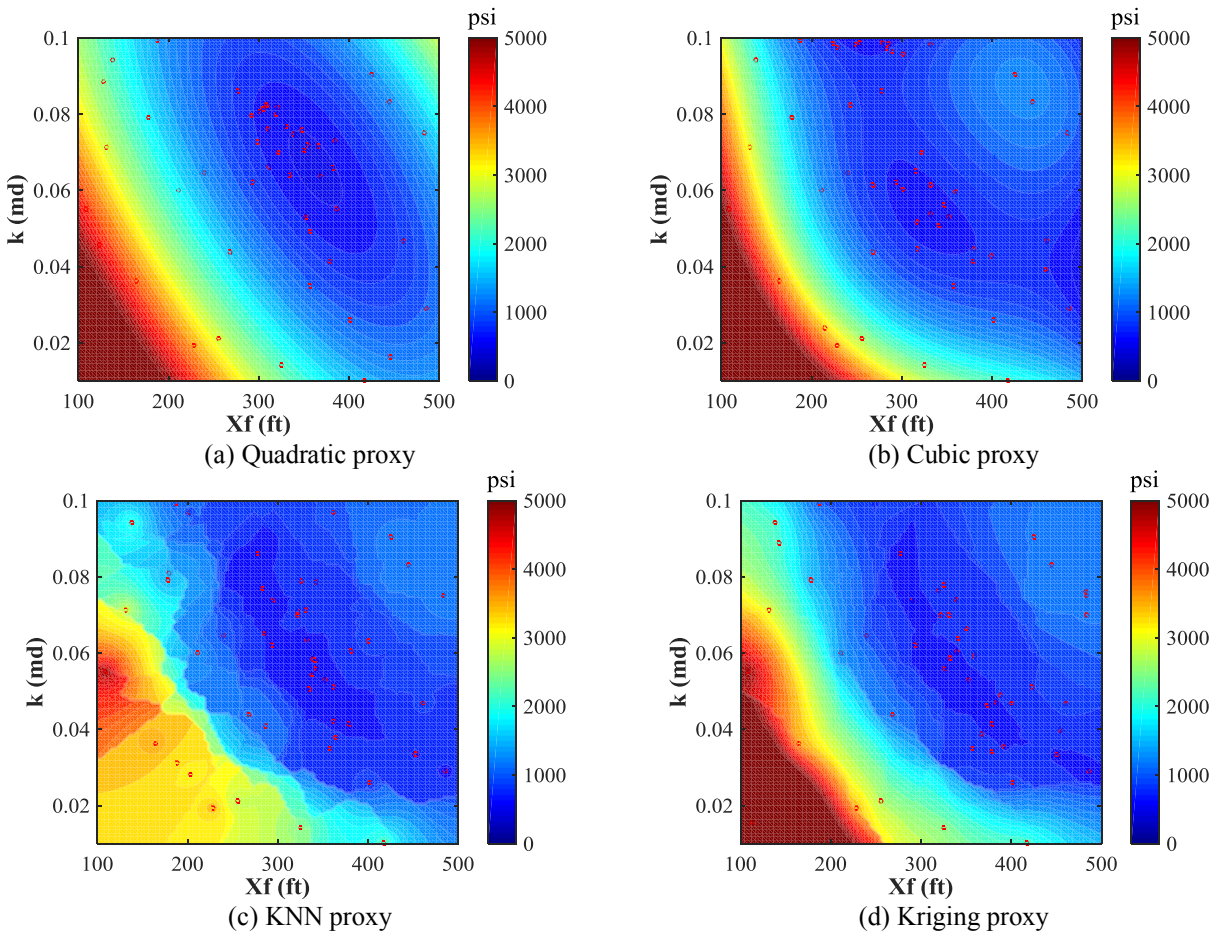
Figure 3.18: Workflow of the history matching study to compare the performances of the different proxy types under different measurement errors



(a) Quadratic proxy (b) Cubic proxy

(c) KNN proxy (d) Kriging proxy

Figure 3.19: Contour plots of the proxy BHP RMSE response surfaces and the corresponding cumulative simulation points for the big-error case at 1st iteration



(a) Quadratic proxy

(b) Cubic proxy

(c) KNN proxy

(d) Kriging proxy

Figure 3.20: Contour plots of the proxy BHP RMSE response surfaces and the corresponding cumulative simulation points for the big-error case at 2nd iteration

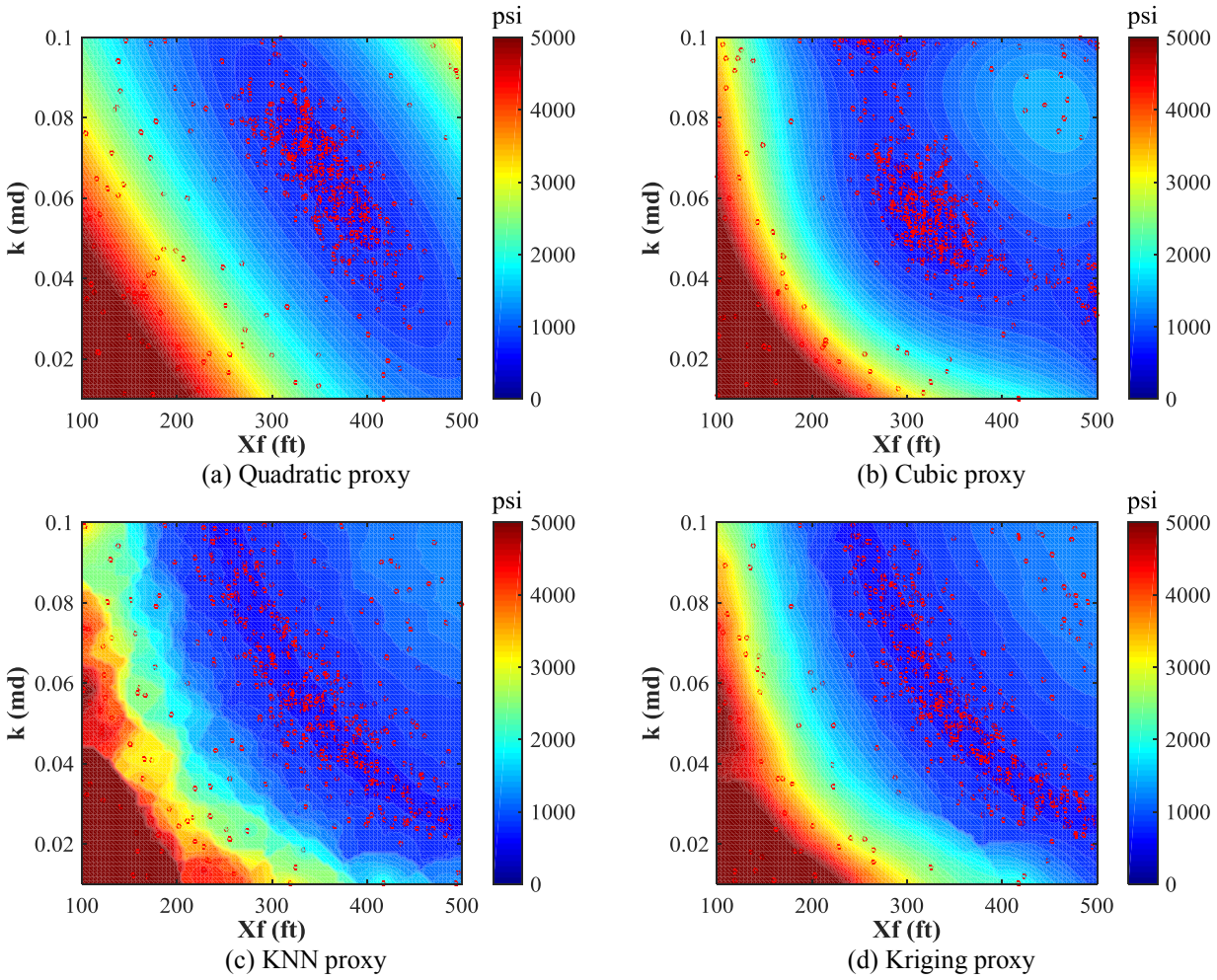
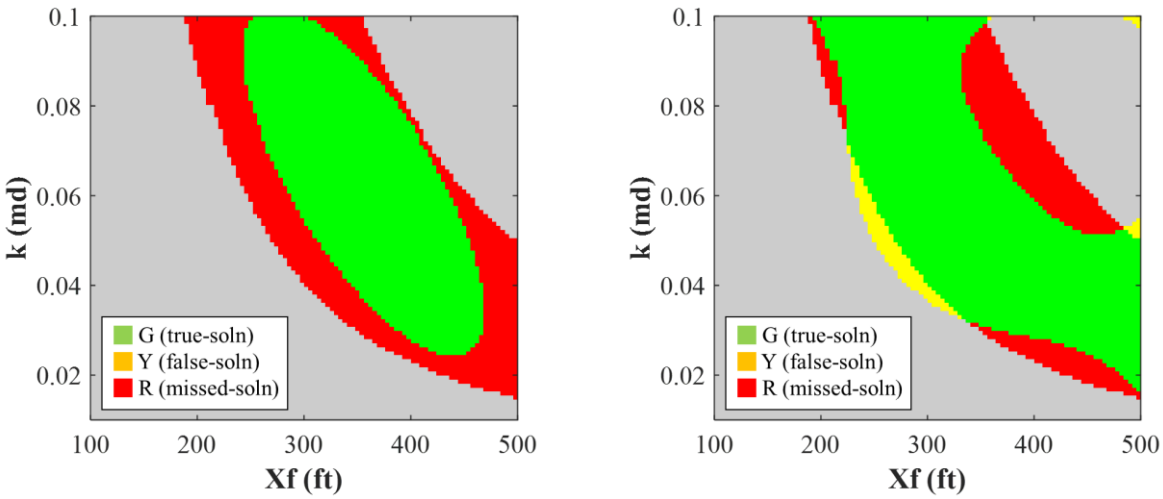


Figure 3.21: Contour plots of the proxy BHP RMSE response surfaces and the corresponding cumulative simulation points for the big-error case at 20th iteration

As can be seen in Figure 3.22, the kriging and KNN proxies replicate the true response surface accurately while the cubic quadratic proxies are less accurate. Figure 3.23 quantitatively compares the AOP of each proxy and reveals that, at the 20th iteration, the most and the least accurate proxies are the kriging and quadratic proxies, respectively. The AOP of the KNN proxy is very close to the cubic proxy, lower than the kriging proxy, but higher than the quadratic proxy. However, KNN requires about 20 iterations to have the

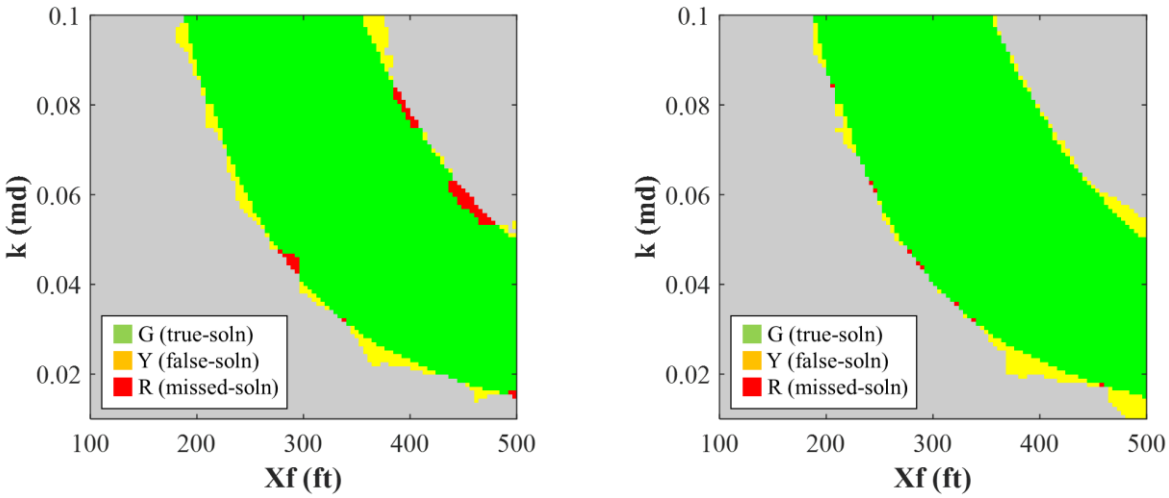
accuracy up to 78% whereas the cubic proxy has the AOP of 79% since the first iteration. If the AOP of 78% is acceptable for this history-matching problem, the cubic proxy is considered more efficient than the KNN proxy. Another interesting observation is that the quadratic and cubic proxies tends to be better at describing the RMSE response surfaces in the case with higher measurement errors. For instance, the AOP of the quadratic proxy at the 20th iteration is 1.7% in the case without measurement error but the AOP is 61.9% in the big-error case. This impact of measurement error may partially explains why quadratic polynomial could have been widely used in the oil and gas industry as an approximately acceptable proxy.



(a) Quadratic proxy

(b) Cubic proxy

Figure 3.22: continued next page.



(c) KNN proxy (d) Kriging proxy

Figure 3.22: The accuracy of proxy maps at 20th iteration of the big-error case (soln in the plots represent solution)

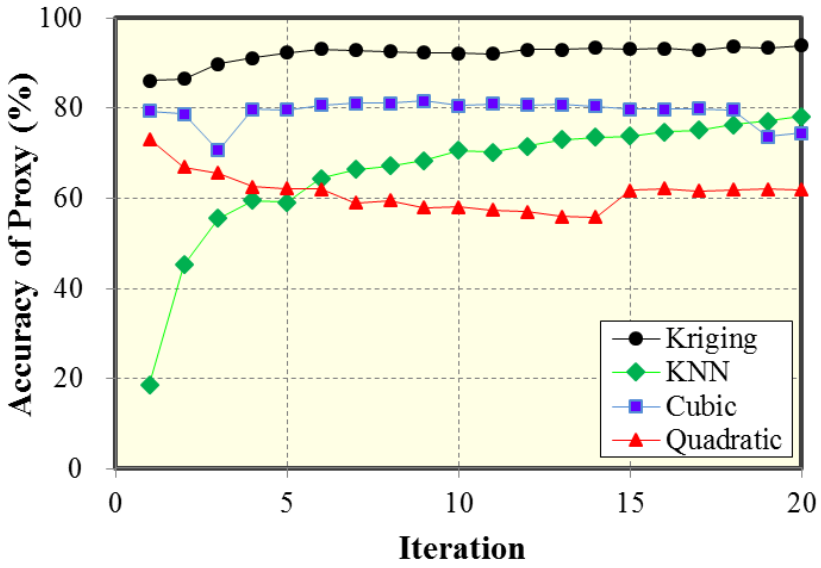


Figure 3.23: Accuracy of proxy using the different proxy types vs iteration for the big-error case

The general recommendations we draw from these observation are as followings. If the computational cost of constructing proxy is not a major concern, kriging proxy is recommended due to its highest accuracy in all the situations we evaluate. Otherwise, we should consider the cheaper proxies such as KNN, cubic, or quadratic proxies. If the response surface is known to be highly complex or the accuracy target is very strict, an exact proxy such as kriging or KNN proxy is recommended. Only if the response surface is known to be simple and the accuracy target can be compromised, an inexact proxy such as cubic or quadratic proxy then may be considered as an acceptable proxy.

3.2 SYNTHETIC FIELD CASE

Incorporating the findings from the synthetic case studies, we implement the AHM workflow on a synthetic field case using a numerical reservoir simulator. The problem description is a horizontal well that is completed with 10-stage hydraulic fracturing in a rectangular dual-porosity reservoir. There are seven uncertain reservoir and fracture parameters as listed in Table 3.2. The historical gas, water flow rates and bottomhole pressure profiles, in Figure 3.24, are synthesized from the historical oil rate profile and the reference parameters to synthesize the history are listed in the last column of Table 3.2. The benefit of using synthetic history is that at least one solution, which is the set of reference values, is known. If the posterior PDFs does not include this solution, the result of the workflow must be invalid and the workflow probably needs longer iterations or other improvements.

Table 3.2: Summary of reservoir and fracture parameters for the synthetic simple case

Parameter	Unit	Minimum	Maximum	Reference
Matrix permeability	nD	10	1000	843
Water saturation	-	0.1	0.5	0.108
Fracture spacing	ft	100	350	307
Fracture half-length	ft	100	500	101
Fracture permeability	D	1	50	11.6
k_{row} exponent	-	2	8	7.44
k_{rw} end point	-	0.1	0.3	0.131

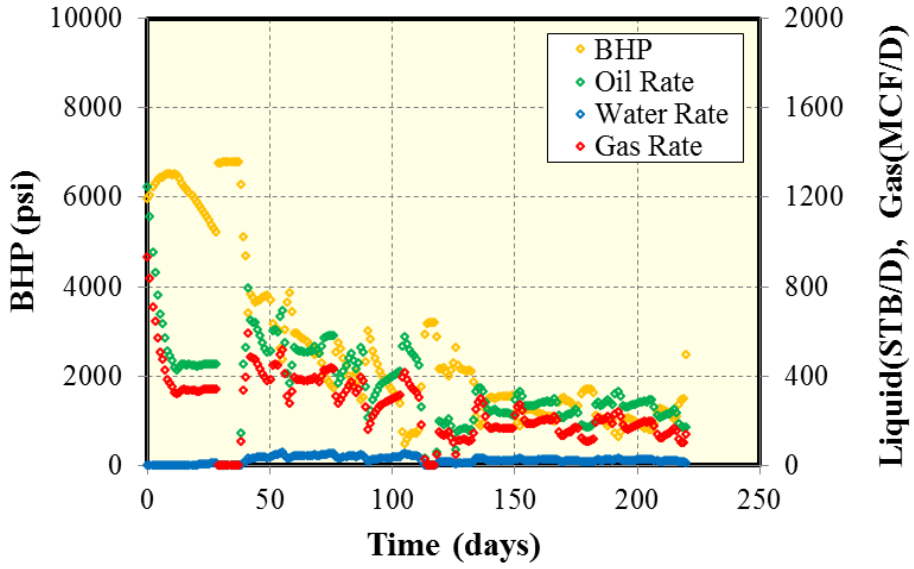


Figure 3.24: Synthetic oil, gas, and water flow rates and bottomhole pressure data

The methodology for the synthetic field case is summarized as the flowchart in Figure 3.25. We use the historical oil rate profile as an input for the reservoir simulator and apply the AHM workflow to history match the three objective functions: BHP, gas rate, and water rate. The proxy chosen is KNN because its efficiency is appropriate for our problem. Before history matching, we run two-level full factorial sensitivity of the seven uncertain parameters and plot the prior 128 simulated results in Figure 3.26. After history matching, we generate 10,000 solution cases and also overlay their posterior simulated

results in Figure 3.26. The comparison of the prior and the posterior PDFs of the uncertain parameters is plotted in Figure 3.27.

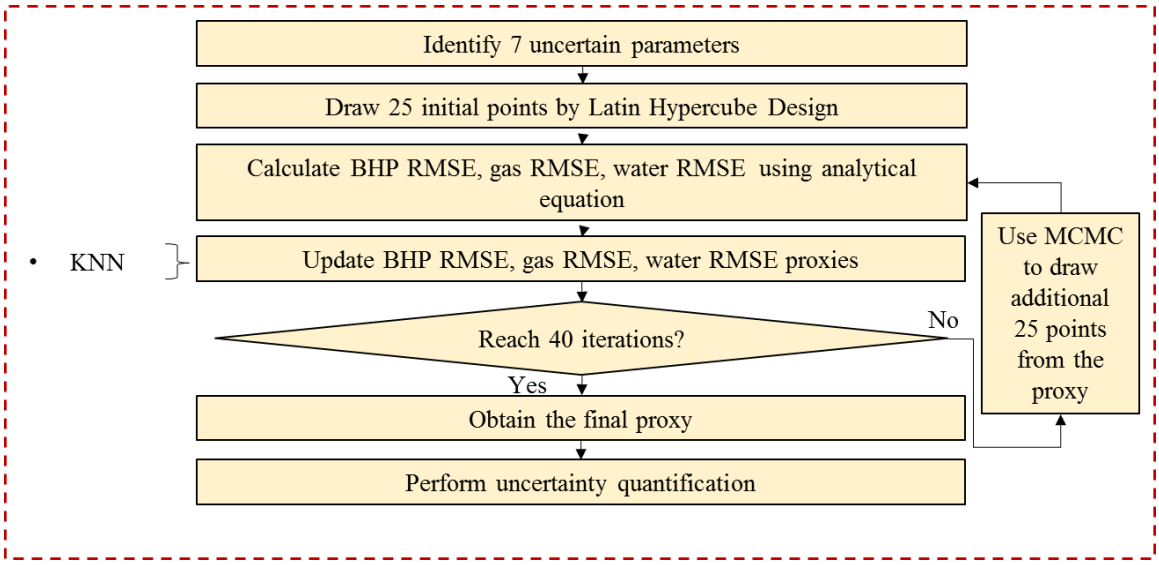
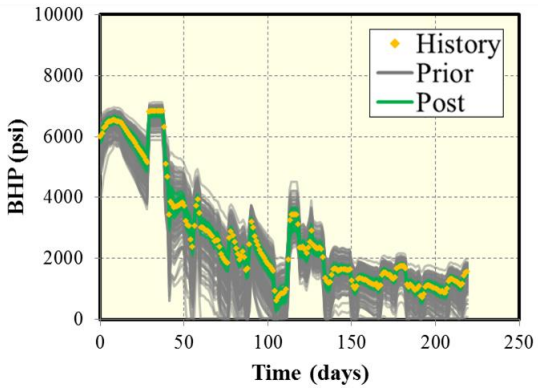
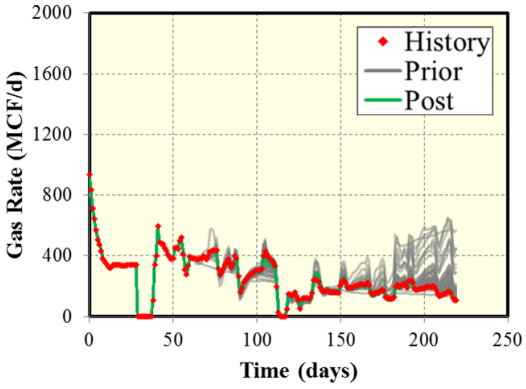


Figure 3.25: Synthetic oil, gas, and water flow rates and bottomhole pressure data

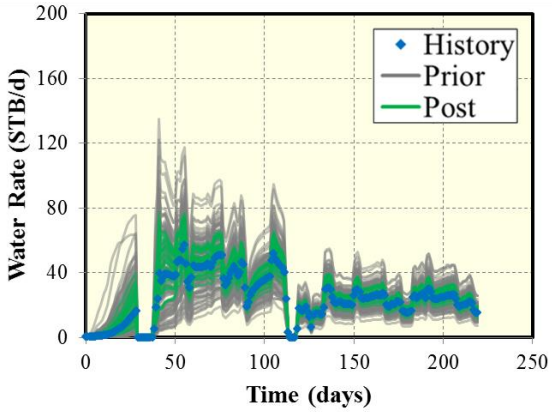


(a) Historical, prior, and posterior bottomhole pressure profiles



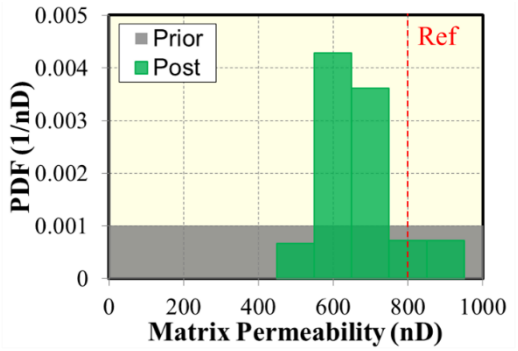
(b) Historical, prior, and posterior gas rate profiles

Figure 3.26: continued next page.

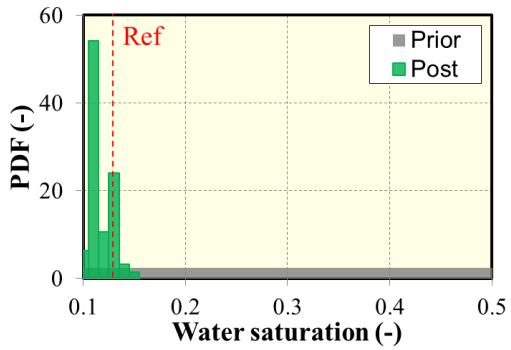


(c) Historical, prior and, posterior water rate profiles

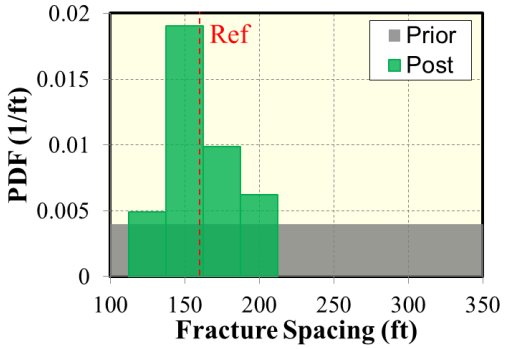
Figure 3.26: Comparison between prior and posterior simulated results



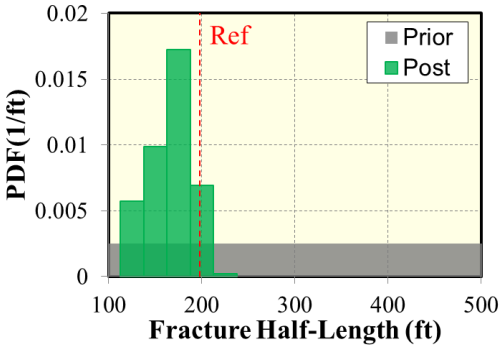
(a) Matrix permeabililty



(b) Water saturation

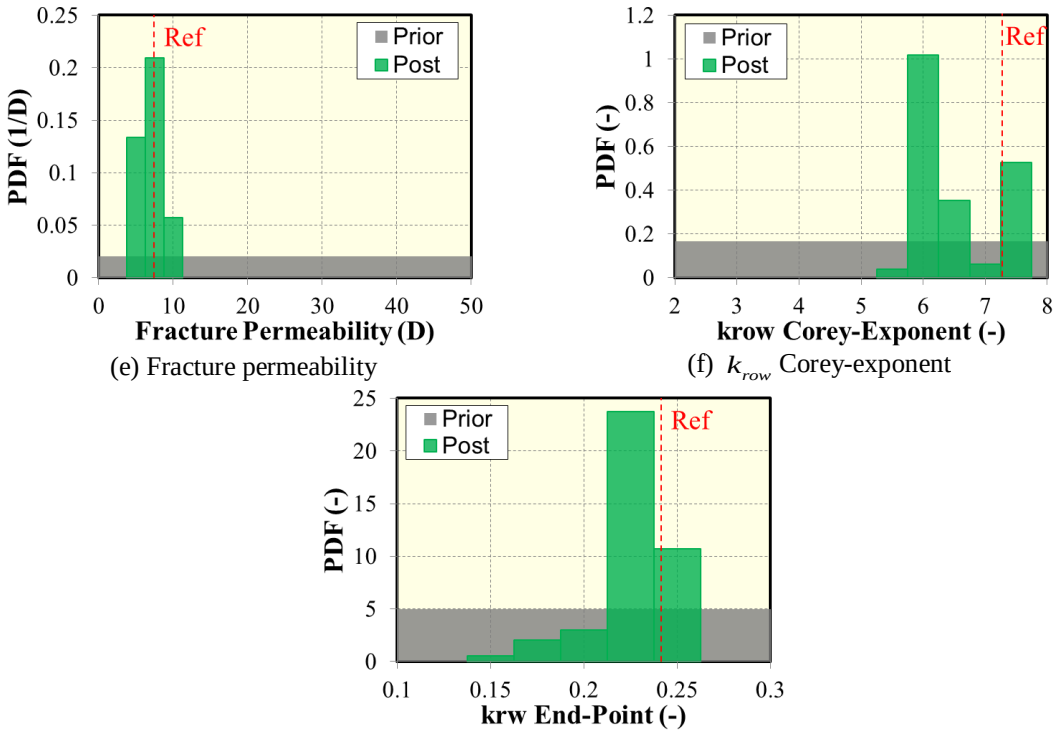


(c) Fracture spacing



(d) Fracture half-length

Figure 3.27: continued next page.

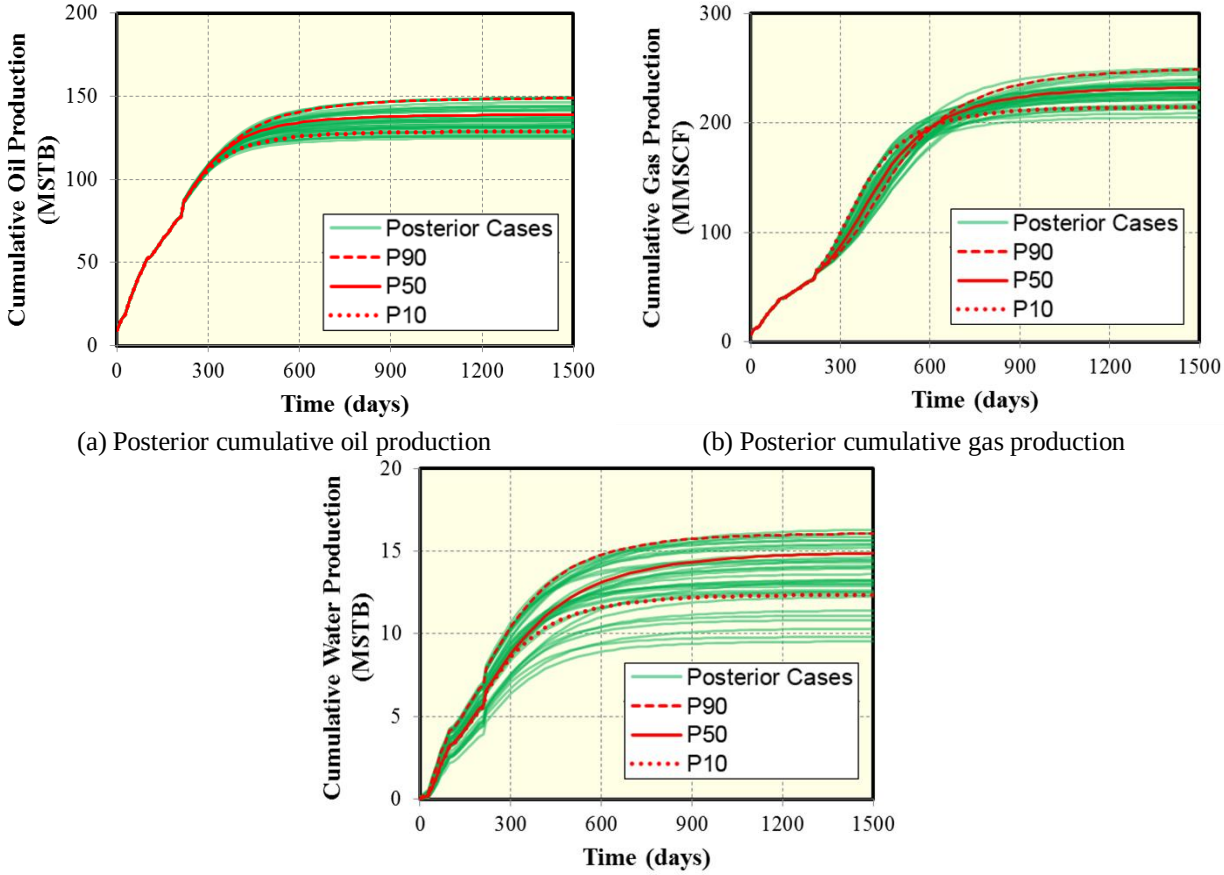


(e) Fracture permeability

(f) k_{row} Corey-exponent

Figure 3.27: Comparison between prior and posterior probability density function of the seven uncertain parameters

For the forecasting in the uncertain quantification phase, the well is assumed to produce at constant BHP of 500 psi from the end of production history to 1500 days. The oil, gas, and water forecasting after history matching are displayed in Figure 3.28. The red lines represent the P10-P50-P90 values and the green lines represent all the 10000 posterior cases. Note that several of the green lines are repetitive due to the MCMC sampling algorithm. We can see in Figure 3.28(a), (b), and (c) that the P10-P50-P90 cumulative production of oil is 129-139-149 MSTB, that of gas is 214-232-249 MMSCF, and that of water is 12.4-14.9-16.1 MSTB. The P90/P10 ratios of all the three cumulative productions are less than 1.3, which indicates the uncertainties are narrower.



(a) Posterior cumulative oil production

(b) Posterior cumulative gas production

(c) Posterior cumulative water production

Figure 3.28: Comparison between prior and posterior probability density function of the seven uncertain parameters

3.3 DISCUSSIONS

There are several advantages of the AHM workflow using proxy-based MCMC. Firstly, the workflow is capable of history matching a complex case like a fractured well in a tight oil reservoir, as demonstrated in the synthetic field case. Secondly, the posterior PDFs using MCMC are statistically valid while some other methods such as optimization-based methods may not provide the accurate PDFs. Thirdly, using proxy can greatly reduce the computational cost that is the main drawback of MCMC. It is useful to note that the choice of proxy type is very important for this purpose. According to our study, polynomial

proxy is not suitable for the strict accuracy target but it may be adequate in the case that the accuracy can be compromised. We recommend using exact proxies such as KNN or kriging to warrant the higher accuracy. Although kriging is slightly more accurate than KNN, we found KNN is computationally more efficient. For example, using KNN proxy requires 0.25 hour as opposed to 8.3 hours using kriging proxy to complete the AHM workflow in one of our cases. For this reason, KNN is the most practical choice in our study.

Some limitations of the AHM workflow are as followings. Firstly, MCMC tends to require more numerical simulations to find the posterior PDFs than optimization-based methods. Secondly, the number of initial simulation points greatly impacts the efficiency of the AHM workflow. Our analysis convinces that too low or too high number of the initial points will result in suboptimal efficiency. However, we do not reach the solid conclusion for the optimal number of the points.

3.4 CONCLUSIONS

The investigation of the AHM workflow using the synthetic simple case discovers several insights for this black-box process. These findings lead to the improvements as well as the new areas to study for the AHM workflow. The key findings are summarized below:

1. Kriging proxy provides more accurate proxy than KNN and cubic proxies. The least accurate is quadratic proxy. In terms of the computational time, however, quadratic proxy is the cheapest, cubic proxy is the second, KNN proxy is the third, and kriging proxy is the fourth. In our experience, KNN proxy is an acceptable proxy providing slightly less accuracy than kriging proxy but requires substantially less computational time.

2. Measurement errors in the observed data smear the precision of the RMSE objective function. In turn, the lower precision reduces the urgency of using very accurate proxy like kriging. In some cases, high measurement errors might even allow a rough model like quadratic to adequately describe the RMSE response surfaces.
3. Although the accuracy of proxy cannot be explicitly calculated in real-world situations, we may use R^2, between the proxy responses of the current iteration and the actual responses of the following iteration, as the guideline to decide the termination of the AHM workflow. The R^2 guideline is appropriate for KNN or kriging proxies but not recommended for quadratic or cubic proxies
4. The number of simulation points at the initial iteration impacts on the AHM workflow efficiency. The low or high number of the points results in the suboptimal efficiency. The guideline for the optimal number is not yet established and will be investigated in our future study.

Chapter 4: Application of the AHM Workflow on Shale Oil Field Case

In this chapter, we apply the AHM workflow using proxy-based MCMC on the real shale oil field case to demonstrate the application. Design of Experiment (DoE) is used as the screening tool at the beginning of the workflow to reduce the number of uncertain parameters. Then the iterative process of proxy-based MCMC is implemented to derive the PPD before probabilistic forecasting is made. We also use the CMOST software (Designed Exploration and Controlled Evaluation algorithm) and the direct MCMC to history match the same case to compare the results with the first workflow in terms of accuracy and efficiency. The lessons learnt of the history matching and the comparative studies are captured.

4.1 Field Case Study in Shale Oil Reservoir

The field case is a shale oil well in Vaca Muerta Field, Argentina. The reservoir is relatively flat with a rectangular shape and approximately uniform thickness. The horizontal well was drilled and completed with several stages of hydraulic fractures which are modelled as 53 bi-wing fractures. The reservoir model, displayed in Figure 4.1, is built using black oil simulation and local grid refinement is used to represent the hydraulic fractures. The 15 months historical data of flowing bottomhole pressure (BHP), oil, and water profiles are available and are plotted in Figure 4.2. The gas rate is however not measured but it is considered negligible. We use the historical oil rate profile as an entry value for the reservoir simulator and aim to history match the BHP and water profiles. The given reservoir and fracture parameters are listed in Table 4.1. The matrix and fracture permeability are expected to reduce when the reservoir pressure decreases. The permeability and pressure relationship is governed by the permeability compliance parameter and Eq. 4.1.

$$k(p) = k_{init} e^{\gamma(p_{init} - p)}, \tag{4.1}$$

where *k(p)* is the permeability at the reservoir pressure , k_{init} is the initial permeability, p_{init} is the initial reservoir pressure, and γ is the permeability compliance.

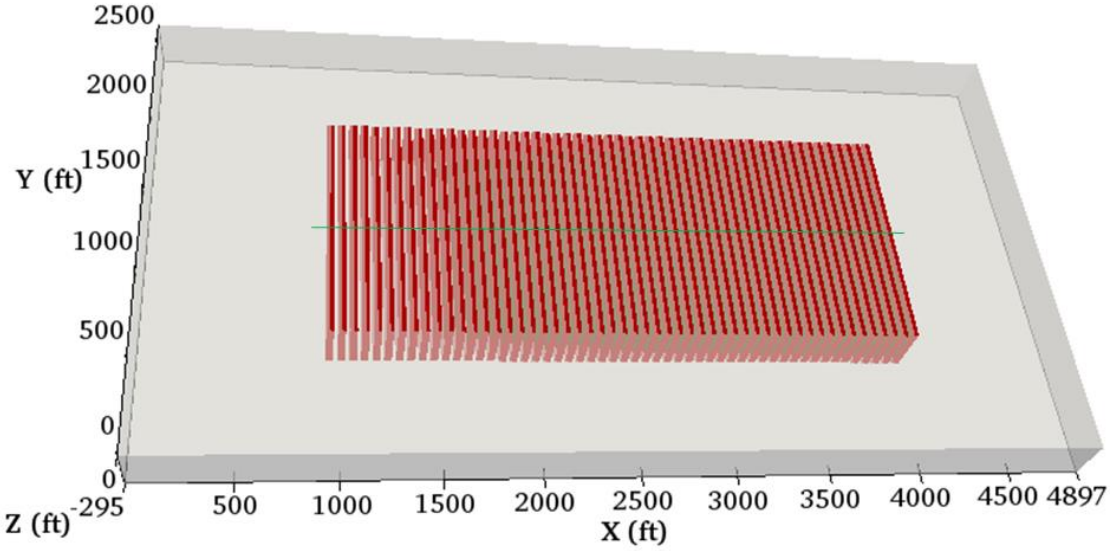


Figure 4.1: Schematic of the basic reservoir model for field case study in Vaca Muerta Formation, Argentina. The red planes represent 53 hydraulic fractures and the green line represents horizontal wellbore

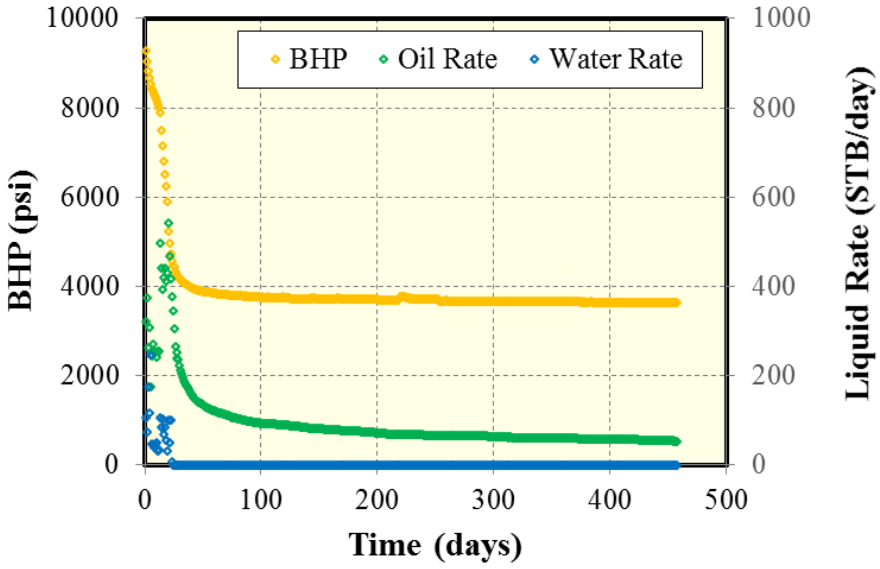


Figure 4.2: Historical oil rate, water rate, and bottomhole pressure data

Table 4.1: Summary of given reservoir and fracture parameters.

Parameter	Unit	Value
Reservoir pressure	psi	9500
Reservoir datum depth	ft	10000
Reservoir length	ft	4897
Reservoir width	ft	2640
Number of fracture	-	53
Fracture spacing	ft	59

4.2 METHODOLOGY

The methodology specific to the field case study is described in the flowchart in Figure 4.3. The first step is to identify the uncertain parameters that possibly affect history matching responses. According to the prior knowledge, eight uncertain parameters are identified and their distributions are assumed to be uniform. A summary of the prior distributions is provided in Table 4.2. In the next step, DOE is used to identify the most influential parameters. The reduction of the number of parameters is helpful since it

dramatically reduces the number of simulation runs required in the subsequent steps. The class of DOE used is two-level factorial design, which can evaluate the effects of individual parameters as well as the effects of the interactions (joint effects changing several parameters simultaneously). The significance of each parameter is quantified by the t-value of its absolute effect. The first parameters with the highest t-values that are responsible for 80% cumulative contribution of significance are considered as the influential parameters.

Table 4.2: Summary of the eight uncertain parameters and their prior distributions

Code	Uncertain parameter	Unit	Distribution	Minimum	Maximum
A	Permeability	mD	Uniform	0.01	0.1
B	Thickness	ft	Uniform	275	315
C	Fracture Half-Length	ft	Uniform	50	650
D	Permeability Compliance	1/psi	Uniform	0	0.0009
E	Water Saturation	-	Uniform	0.2	0.5
F	Porosity	-	Uniform	0.001	0.08
G	Formation Compressibility	1/psi	Uniform	5.8E-6	9E-6
H	Fracture Conductivity	md-ft	Uniform	100	1000

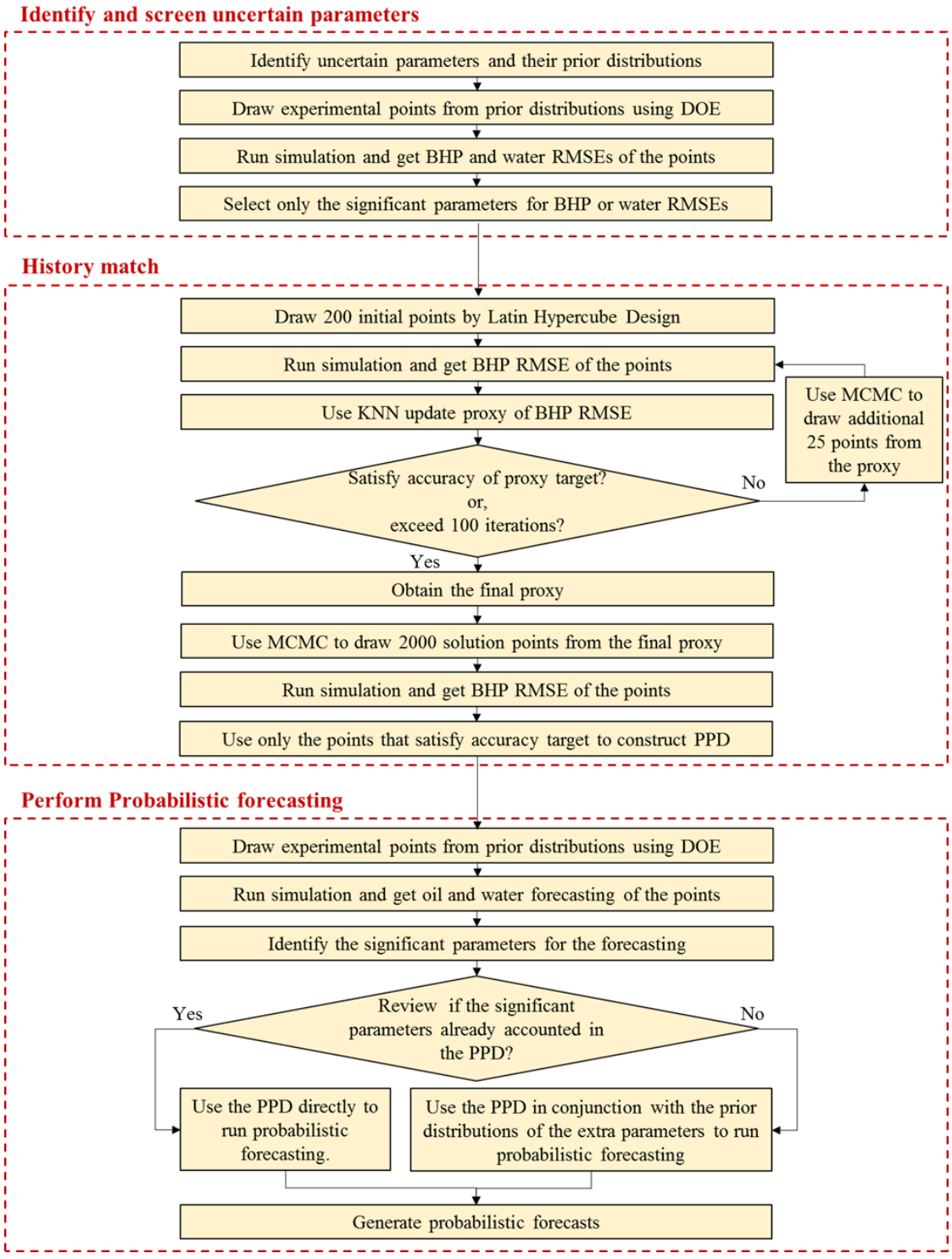


Figure 4.3: Flowchart of the AHM workflow using proxy-based MCMC applied on the shale oil field case.

Latin Hypercube Design (LHD) is then used to randomly draw 200 initial points from the domain of the influential parameters. Next, reservoir simulations are performed to attain BHP RMSE of these 200 points. Based on these simulation points, KNN proxy can be used to approximate BHP RMSE of any other points in the domain. LHD is preferable because the points sampled by this design provide more accurate proxy than other traditional designs such as Placket-Burman, central composite and D-optimal (Yeten et al., 2005). MCMC is then used to sample the entire proxy response surface and construct the PPD. We randomly choose another 25 additional points from the PPD and run reservoir simulations to obtain their BHP RMSEs. In the second iteration, the proxy is updated using all the points available including the initial points and the additional points. After that, MCMC is used to sample the updated proxy and construct the PPD.

$$\mathrm{RMSE}(\vec{\theta}) = \sqrt{\left\{\sum_{j=1}^{n}\left[d_{obs}(t_j) - d_{sim}(\vec{\theta}, t_j)\right]^2\right\} / n}\,, \tag{4.2}$$

The workflow is run iteratively until the maximum limit, set to 100 iterations, is exceeded or until the proxy is sufficiently accurate enough. The accuracy of the proxy is difficult to explicitly measure; hence, we use several metrics that indirectly quantify the accuracy. The metrics include the absolute error, described by Eq. 4.3, and the iteration-wise R^2, described by Eqs. 4.4 and 4.5. If the P90 absolute error is less than 300 psi or the iterative-wise R^2 is higher than 0.8, the accuracy of the proxy is satisfactory.

$$E(P^{th}) = P^{th}\,percentile\left(\left|RMSE_{proxy,j,i} - RMSE_{actual,j+1,i}\right|\right), \tag{4.3}$$

$$R_j^2 = 1 - \frac{\sum_{i=1}^{n} (RMSE_{proxy,j,i} - RMSE_{actual,j+1,i})^2}{\sum_{i=1}^{n} (RMSE_{actual,j+1,i} - \overline{RMSE}_{actual,j+1})^2}, \tag{4.4}$$

$$\overline{RMSE}_{actual,j+1} = \frac{1}{n} \sum_{i=1}^{n} RMSE_{actual,j+1,i}, \tag{4.5}$$

where E is the absolute error at the P^{th} percentile, n is the number of simulation points in an iteration, $RMSE_{proxy,j,i}$ is the proxy RMSE and $RMSE_{actual,j+1,i}$ is the actual RMSE. The subscript j and $j+1$ depict the iteration number. The subscript i depicts the simulation point i out of the total n points in an iteration.

After the termination of the workflow, the final BHP RMSE proxy is constructed. MCMC is then used to thoroughly sample the final proxy and construct the PPD. We randomly pick 2000 points in the PPD with proxy BHP RMSE of lower than 300 psi and run reservoir simulation for these points. Finally, only the points with the BHP RMSEs lower than 300 psi and the simulated water RMSE lower than 20 STB/day are accepted as the history-matched solutions and used to construct the final PPD.

While it seems intuitively correct to use only the history matched realizations for the probabilistic forecasting, it should be noted that a parameter trivial for history matching can be influential for forecasting (Busby et al., 2009). Therefore, all the uncertain parameters must be reconsidered. The two-level factorial design is used to analyze and determine which parameters significant for the forecasting phase. If these significant parameters are already taken into account during the history matching phase, we can directly use the final PPD previously constructed. Otherwise, the uncertainty of the extra parameters must be considered in conjunction with the final PPD.

4.3 RESULTS

During DOE step, 256 reservoir model realizations are generated using a two-level full factorial design for the eight parameters in Table 4.2. The Pareto charts of their BHP RMSEs and water RMSEs are plotted in Figure 4.4 and Figure 4.5, respectively. The definition of each term on the x-axes is consistent to the code in Table 4.2. As Figure 4.4 shows, the terms DF, F, AEF, ADEF, ADE, ACDF, AE, and CF cumulatively contribute more than 80% of the effect on BHP RSMSE. For the water RMSE, the terms E, D, DE, DEF, DF, C, CE, and F are the major contributors, which can be seen in Figure 4.5 Therefore, the A, C, D, E, and F parameters (permeability, fracture half-length, permeability compliance, water saturation, and porosity) are the most significant parameters responsible for the BHP and the water RMSEs. These five parameters will be varied while the other three parameters are kept constant at reference values in the further steps.

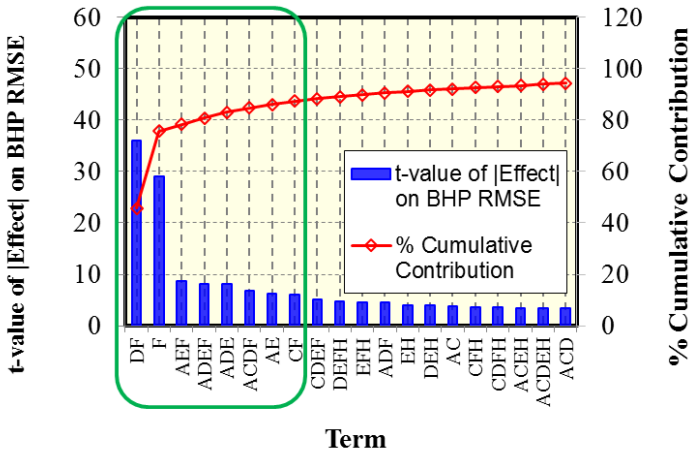


Figure 4.4: Pareto charts of BHP RMSE according to the two-level full factorial design.

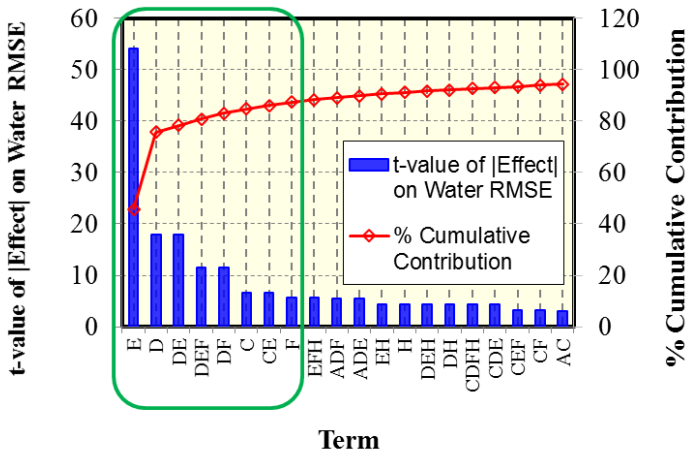
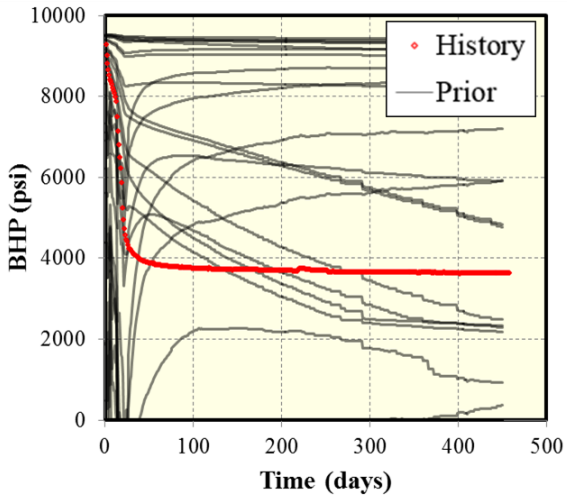
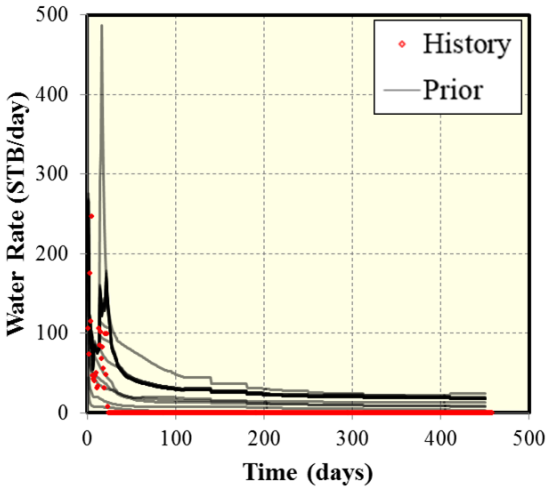


Figure 4.5: Pareto charts of water RMSE according to the two-level full factorial design.

In addition to the workflow in Figure 4.3, we plot the forecasts based on a full factorial design of the five significant parameters, for the total of 32 cases, as displayed in Figure 4.6. This plot helps visualize the diversity of the possible forecasts. As can be noticed, the simulated BHP profiles in Figure 4.6(a) are significantly diverse; the pressure profile remains nearly flat in a case while the pressure of another case abruptly drops to zero. The vast difference of the BHP responses indicates the high uncertainty in the prior distributions. Thus, a high computational cost to achieve a proper history matching is expected. In some situations, reservoir engineers can use other information to narrow down the prior distribution before proceeding to the next step. In this problem, however, we accept the prior distributions due to unavailability of the extra information.



(a) Historical and simulated bottomhole pressure profiles



(b) Historical and simulated water rate profiles

Figure 4.6: Historical profiles and the possible 32 simulated profiles based on the prior distribution of the five significant parameter using full factorial design

The workflow is terminated because the limit of 100 iterations is exceeded but not because the accuracy targets are satisfied. The iteration-wise R^2 is calculated and plotted in Figure 4.7 . As can be seen, the trend of the R^2 at the 100th iteration is still lower than the target of 0.8. Likewise, the absolute error, shown in Figure 4.8, indicates that the P90 absolute error at the 100th iteration is still higher than the target of 300 psi. It is common that computational time or number of simulations are the termination criteria for large problems (Landa et al., 2003). The high uncertainty in the prior distributions is likely a major cause of the complexity in this case study.

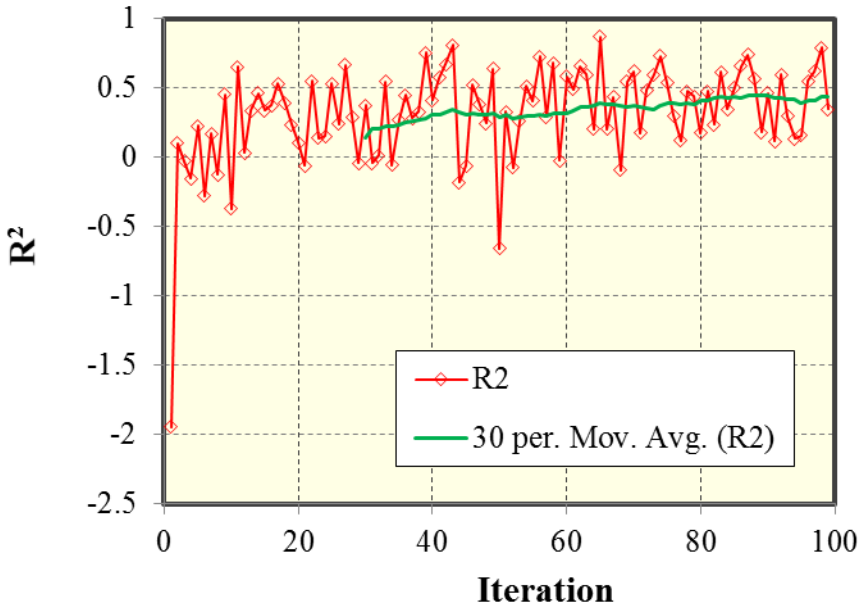


Figure 4.7: The change of R^2 with the iteration process. The moving average of R^2 is approximately 0.4 after the 100th iteration

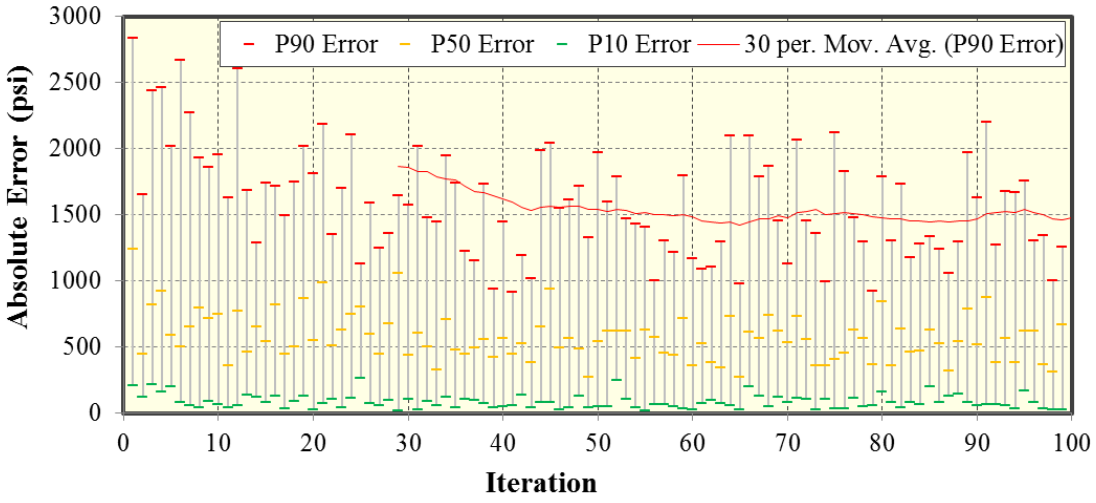


Figure 4.8: P10, P50, and P90 absolute error between the proxy RMSE and the actual RMSE during the iteration process

There are a few ways to improve the efficiency of this history matching task. The first way involves with narrowing down the uncertainty ranges. In the view of many researchers (Slotte et al., 2008; Busby et al, 2009), one significant step in the history matching process is to define appropriate uncertain parameters and prior distributions. If an unreasonable prior knowledge is assumed, the resultant PPD may be invalid regardless of robust AHM workflows used. The revision of the prior distribution can be made by using extra information or consulting with the right experts. The second way is to change the proxy type from KNN to a more accurate proxy, such as kriging. However, using kriging poses another drawback with a high computational cost. While kriging proxy is expected to reduce the number of simulations, its computational time is impractically high for the standalone computer used in this study. If a high performance cluster is available, kriging would have been considered.

MCMC is used to sample the final BHP RMSE proxy and suggest the best 2000 realizations. The simulation is performed for these realizations and 52 realizations of them are found to be the actual history-matching solutions (less than 300 psi BHP RMSE and less than 20 STB/day water RMSE). The final solutions are summarized as the parallel coordinate plot shown in Figure 4.9. The plot is a powerful way to visualize the multi-dimensional realizations. Each of the green lines represents one of the history-matched realizations connecting the five parameters in the first five vertical axes. Since the realizations are not equiprobable, their sampling frequencies must be attained from MCMC algorithm and the frequencies are displayed in the sixth axis. To display the history-matching solutions in a traditional view, Figure 4.10 displays the histograms that are the projection of these solutions onto each of the parameters. The histograms are also overlaid and compared with the prior distributions. The stimulated rock volume (SRV) is likely lower than initially expected since the posterior fracture half-length and the porosity are

mostly located at the low sides, as shown in Figure 4.10(b) and Figure 4.10(c), It is important to note that each of the histograms is interdependent, i.e., not necessarily the most likely realization is simultaneously sampled from the mode of every parameter. Another interesting result is the bi-model permeability as can be seen in Figure 4.10(a). This could be a correct distribution or probably an inaccurate one as a consequence of low accuracy of proxy. Figure 4.11 displays the comparison between the production history and simulation. As shown, the posterior profiles of both BHP and water rate have good agreement with the historical profiles.

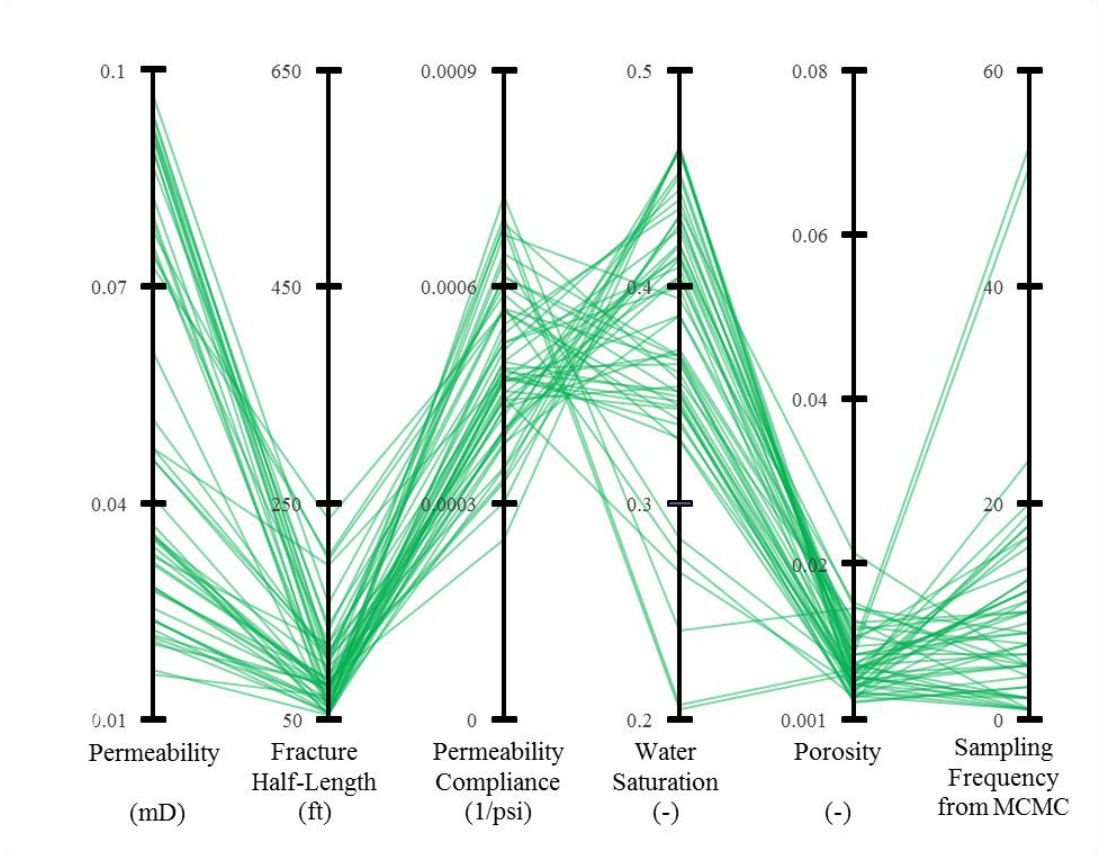
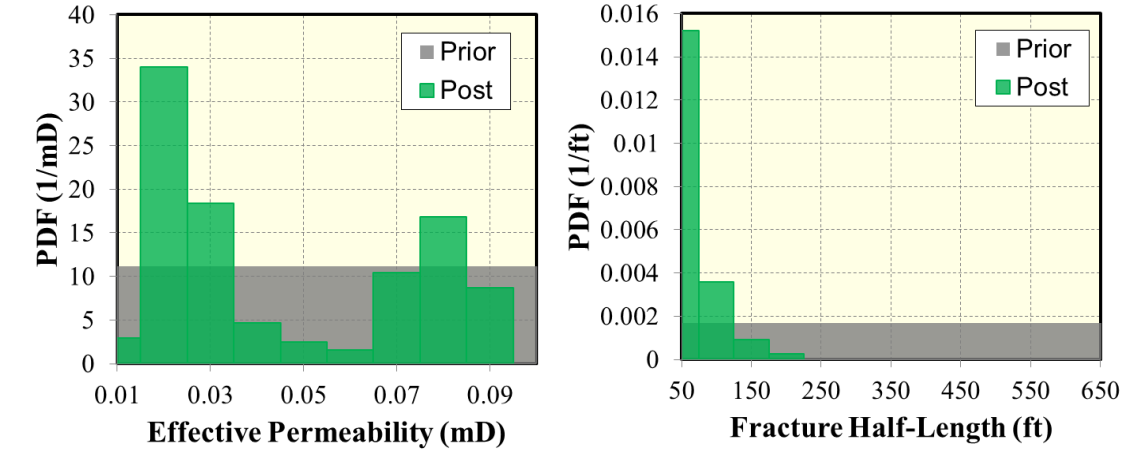
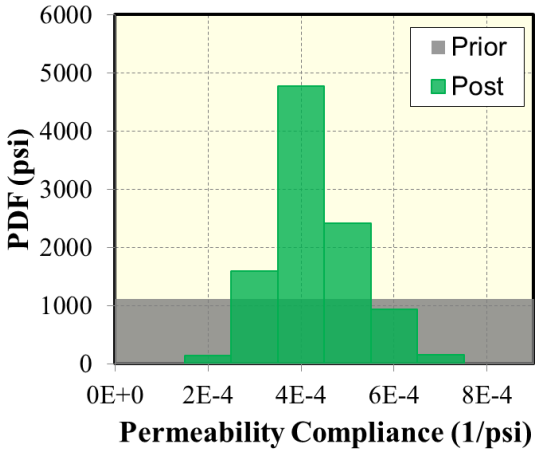


Figure 4.9: Parallel coordinate plot of 52 history-matched solutions found by the proxy-based workflow

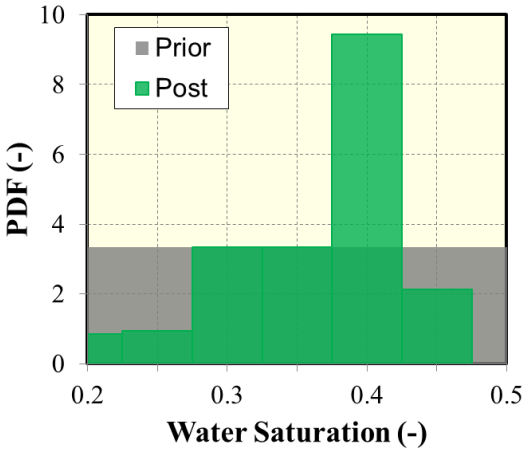


(a) Effective permeability

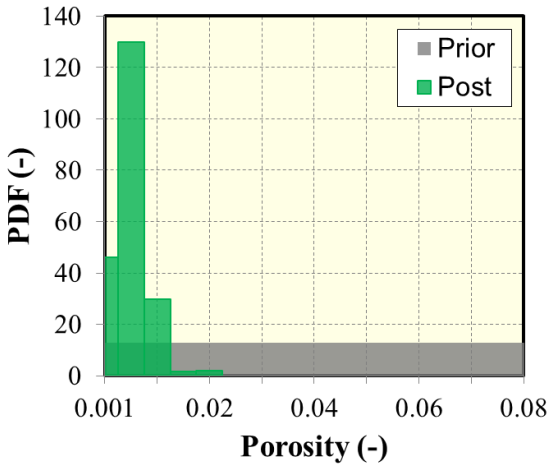
(b) Fracture half-length



(c) Permeability compliance

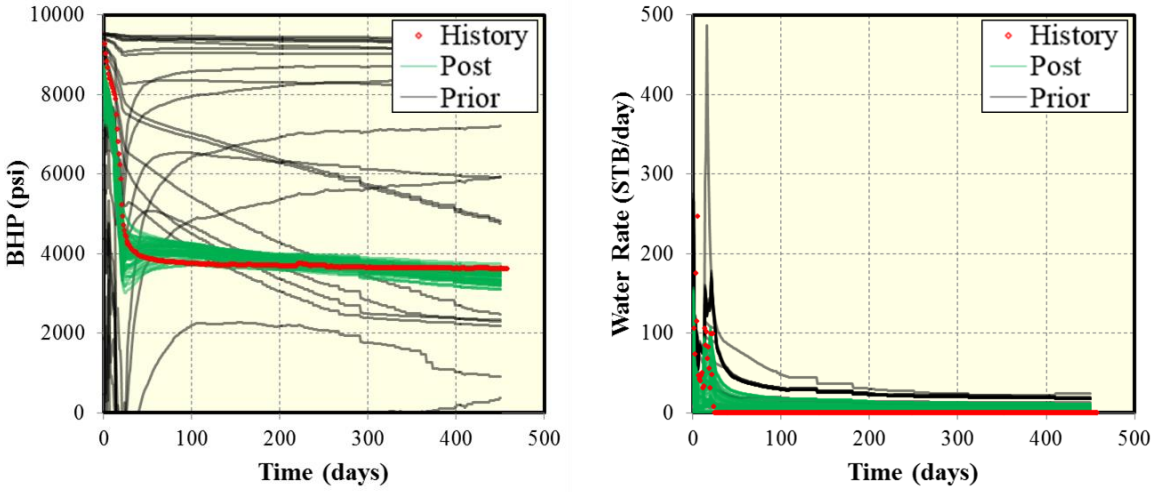


(d) Water saturation



(e) Porosity

Figure 4.10: Comparison between prior and posterior probability density function of the five significant uncertain parameters using the proxy-based MCMC



(a) Historical and simulated bottomhole pressure profiles

(b) Historical and simulated water rate profiles

Figure 4.11: Comparison between prior and posterior simulated results

For the probabilistic forecasting phase, the well is scheduled to produce with BHP of 500 psi until 8000 days. Two-level full factorial design is used to analyze the effects of all the eight parameters, in Table 4.2, on the forecasts of cumulative oil production, cumulative water production, and oil recovery factor. The Pareto charts of the forecasts are presented in Figure 4.12, Figure 4.13, and Figure 4.14, respectively. As can be seen in the green rectangles, the charts reveal that the most influential parameters for the three forecasts are C, D, E, and F. Since these parameters are already considered in the history matching phase, the final PPD previously constructed can be directly used to run the probabilistic forecasting.

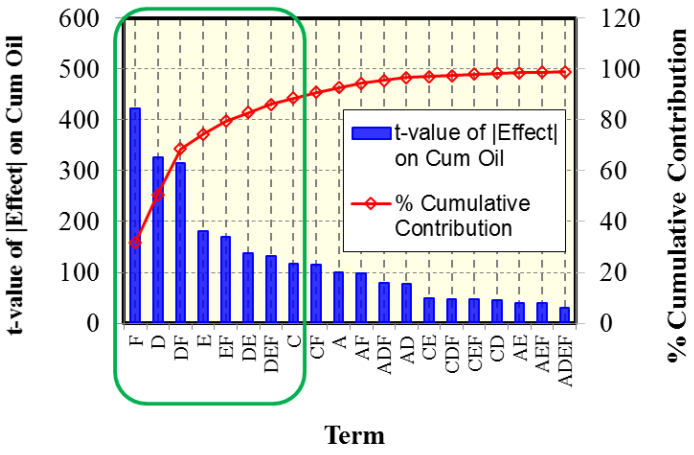


Figure 4.12: Pareto charts of cumulative oil production forecast according to the two-level full factorial design.

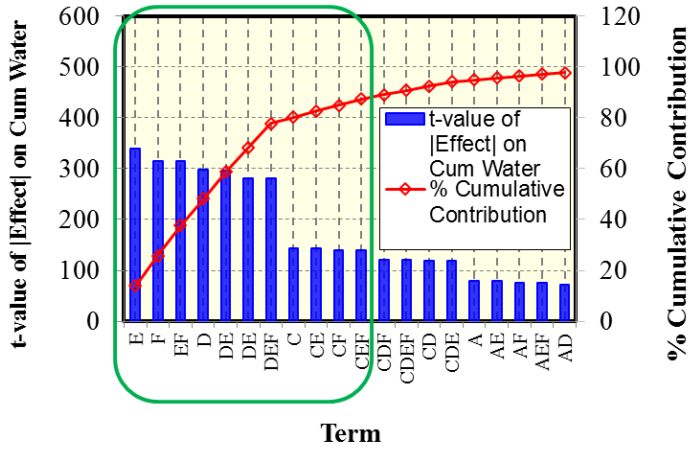


Figure 4.13: Pareto charts of cumulative water production forecast according to the two-level full factorial design.

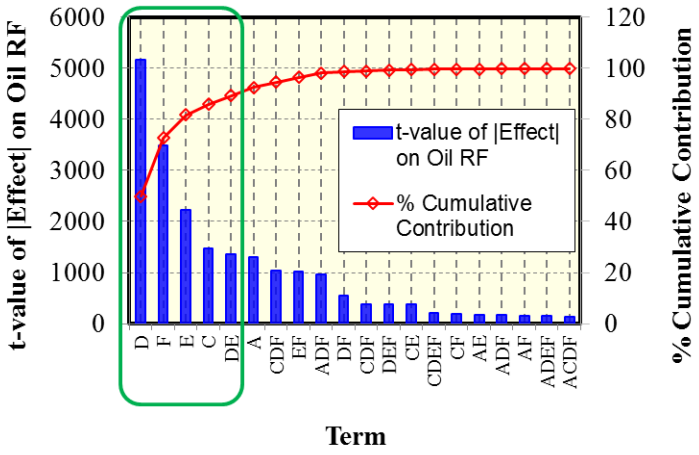
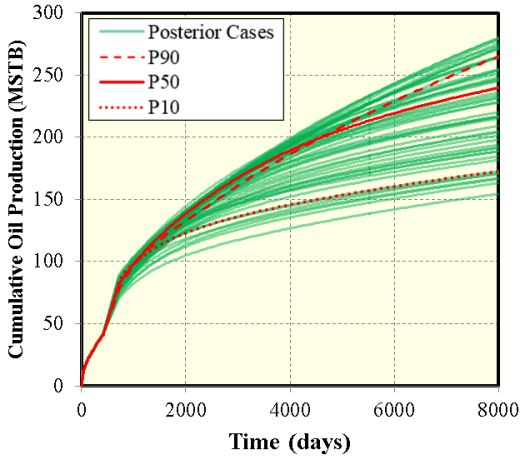
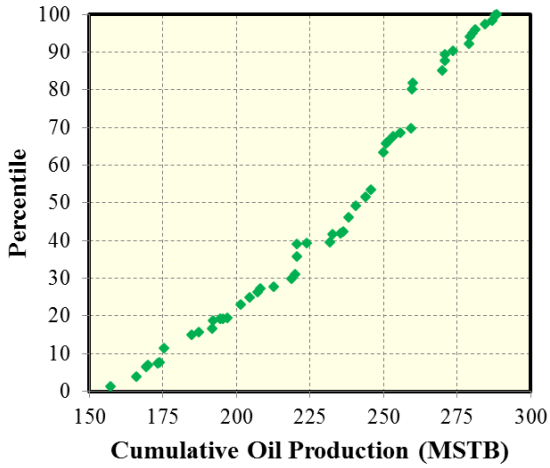


Figure 4.14: Pareto charts of oil recovery factor forecast according to the two-level full factorial design.

The posterior forecasts of cumulative oil production, cumulative gas production and oil recover factor are plotted in Figure 4.15, Figure 4.16, and Figure 4.17, respectively. The probabilistic forecasts are plotted in the left panels where the green lines and the red lines depict the P10, P50, and P90 values. The s-curves of the forecasted values after 8000 days are displayed in the right panels. As shown in Figure 4.15, the P10-50-90 cumulative oil production is 175-244-274 MSTB. The ratio of P90/P10 of 1.6 indicates moderate uncertainty in the oil production forecasts. The uncertainty in the water production forecasts is also appreciable since the P90/P10 ratio is about 5 according to Figure 4.16. The P10-50-90 oil recover factor of the Vaca Muerta well is 8%-10%-14% as shown in Figure 4.17.

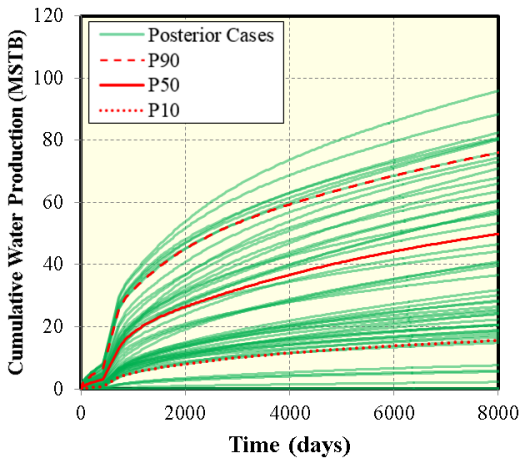


a) Cumulative oil production vs time

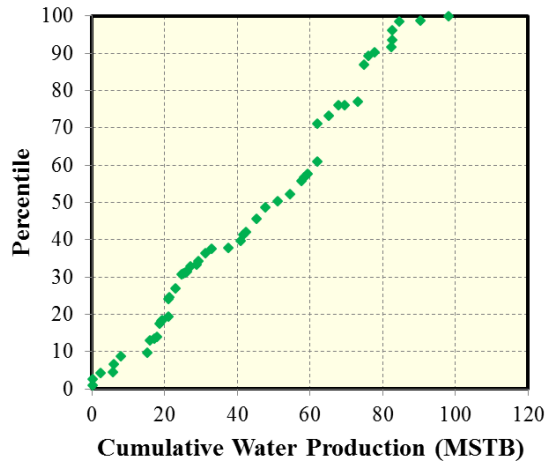


b) S-curve of cumulative oil production after 8000 days

Figure 4.15: Posterior cumulative oil production forecasts.

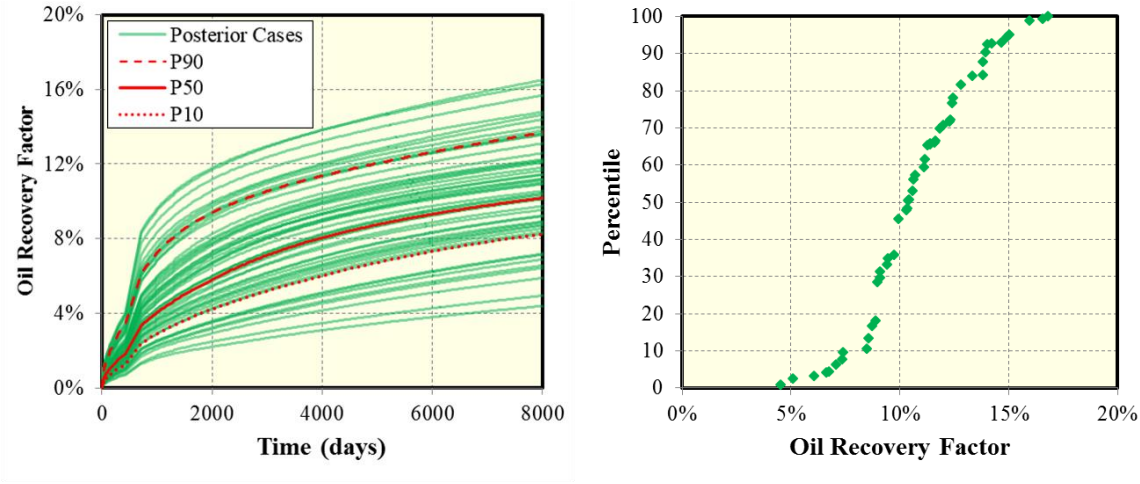


a) Cumulative water production vs time



b) S-curve of cumulative water production after 8000 days

Figure 4.16: Posterior cumulative water production forecasts.



a) Oil recovery factor vs time

b) S-curve of oil recovery factor after 8000 days

Figure 4.17: Posterior oil recover factor forecasts.

4.4 Comparison of The Workflow with A Commercial History Matching Software

We use CMOST software (CMG, 2014) to history match the similar field case study in order to compare the performance with the proxy-based MCMC workflow. The specific algorithm used is DECE since the algorithm is one of the recommended optimization methods in CMOST. DECE is an intellectual property of CMG (2014) and the theory is revealed in detail. In brief, the algorithm is an optimization method that can be described as an iterative optimization process. The algorithm is comprised of a designed exploration stage and then a controlled evolution stage. In the former stage, the goal is to explore the parameter domain in a designed random manner such that maximum information about the solution space can be obtained. Experimental design and Tabu search technique are applied to select parameter values and create realizations of simulation models. In the latter stage, statistical analyses are performed for the simulation results obtained in the designed exploration stage. Pertaining to the analyses, the DECE algorithm scrutinizes every candidate value of each parameter to determine whether the solution quality can be

improved if certain candidate values are rejected from being picked again. These rejected candidate values are remembered by the algorithm and will not be used in the next controlled exploration stage. However, to minimize the possibility of being stuck in local minima, the DECE algorithm checks the rejected candidate value from time to time to make sure previous rejection decisions are still valid. If the algorithm determines that certain rejection decisions are not valid, the decisions are recalled and the corresponding candidate values are used again.

Table 4.3 summarizes the efficiency of CMOST and compares with that of the proxy-based MCMC workflow and CMOST. As can be seen for the similar total simulation runs of 4675, CMOST finds more than ten times the number of history matching solution the proxy-based MCMC workflow does. We also track the efficiency versus the number of simulation for the first 2500 simulation runs, as plotted in Figure 4.18.

Table 4.3: Comparison of the efficiency between the proxy-based MCMC workflow and CMOST

Parameter	Proxy-Based MCMC workflow	CMOST
Simulation runs to improve proxy	2675	-
Simulation runs to explore final proxy	2000	-
Total simulation runs	4675	4675
History matching solutions	52	533
Total simulation runs/solutions	89.9	8.8

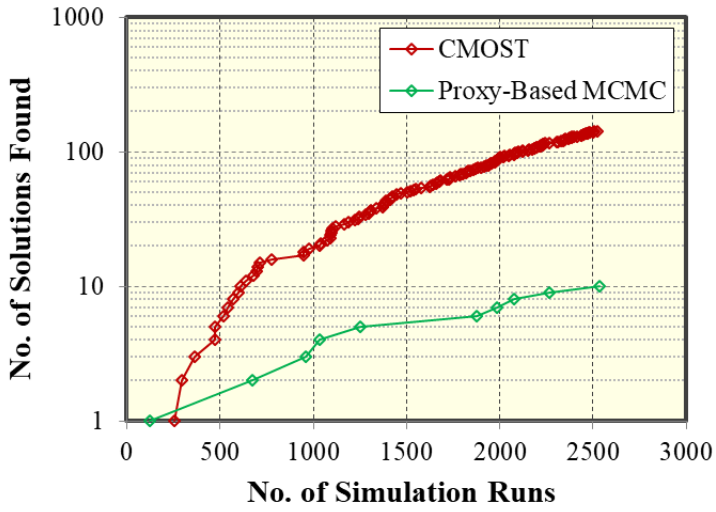


Figure 4.18: Comparison of efficiency to find solutions of the proxy-based MCMC workflow and CMOST.

However, the efficiency is not the only aspect to evaluate the performance of a history matching workflow. The other important aspect is the accuracy. The accurate PPD should cover all the range of possible solutions. Figure 4.19 displays the PPD of the history-matching solutions from CMOST. The first five leftmost axes depict the five significant uncertain parameters. The rightmost axis represents the sampling frequency of each history-matching solution, which each of the frequency equals to one when using CMOST. Comparing this PPD and the PPD by the proxy-based MCMC workflow in Figure 4.9 reveals that CMOST still misses some of the solutions found by the proxy-based MCMC workflow despite the algorithm in CMOST to minimize being stuck in local minima.

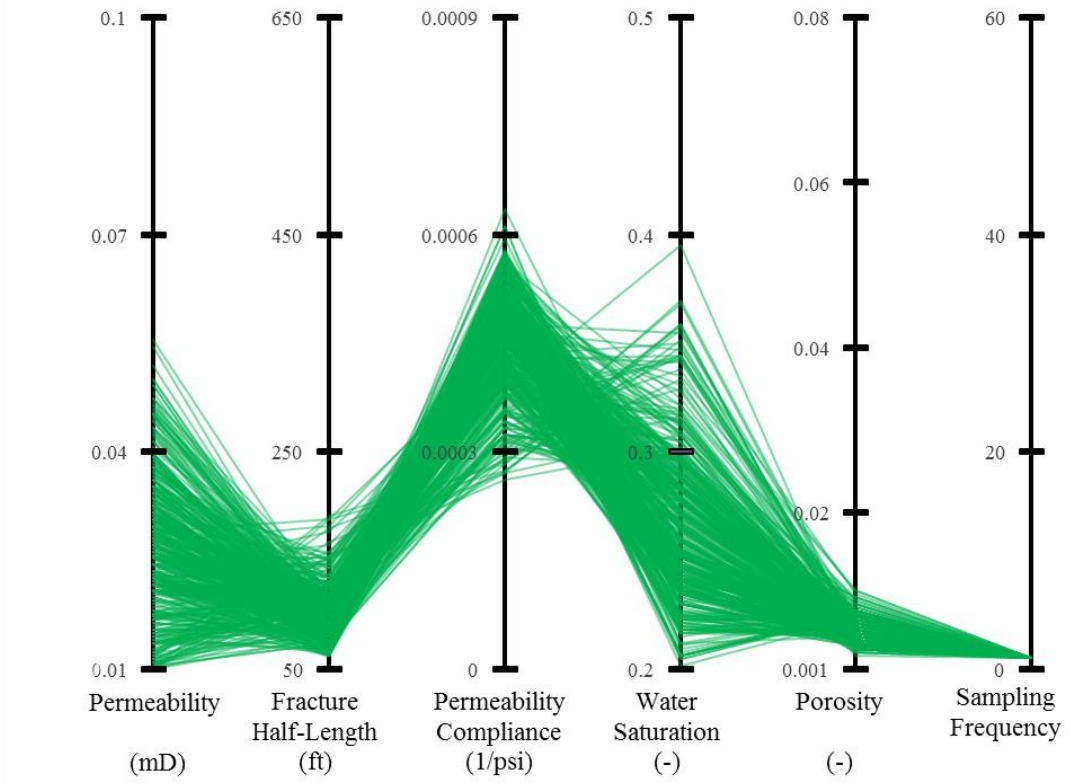
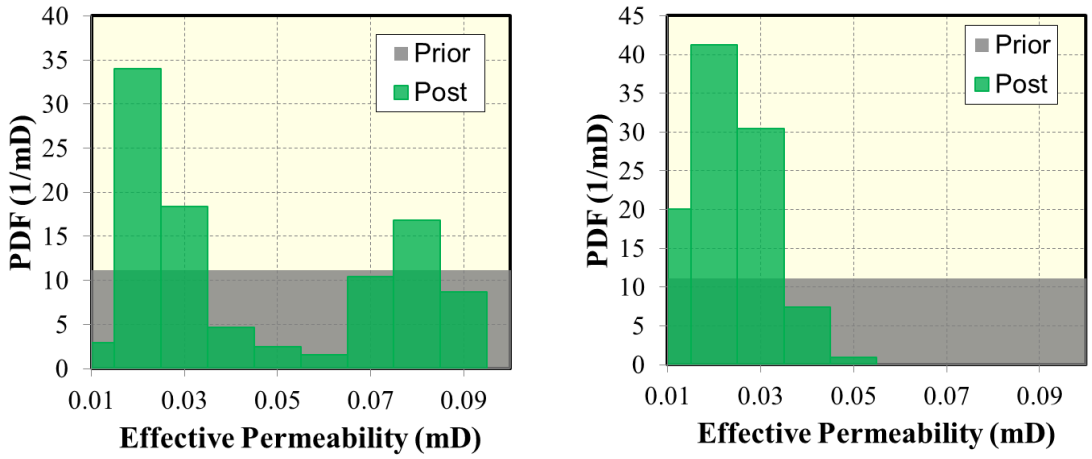


Figure 4.19: Parallel coordinate plot of history-matched solutions found by CMOST software.

The accuracy of the solutions can obviously be noticed when comparing the permeability axes of the two method. The histograms of the permeability axes are also displayed in Figure 4.20. While the history-matched permeability by the proxy-based MCMC workflow ranges from 0.01 to 0.09 mD in Figure 4.19(a), the same by CMOST ranges from 0.01 to 0.05 mD in Figure 4.19(b). To show spreading of the solutions, history-matched permeability is plotted against the number of simulation runs for the two methods in Figure 4.19 where a marker in each line represents a solution found. We can see that after 1000 simulation runs, CMOST finds 18 solutions with the permeability ranging from

0.01 to 0.03 mD. On the other hand, the proxy-based MCMC finds only three solutions but with more diverse permeability, ranging from 0.01 to 0.09 mD.



a) Proxy-based MCMC workflow b) CMOST

Figure 4.20: Comparison of the history-matched permeability histogram between the proxy-based MCMC workflow and CMOST

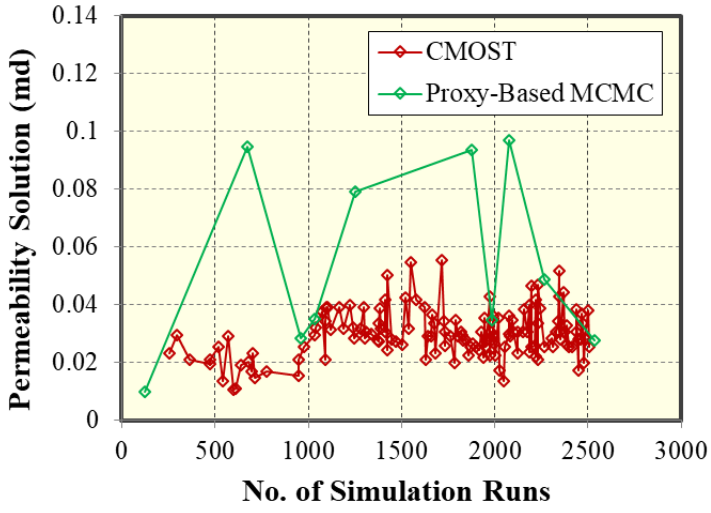
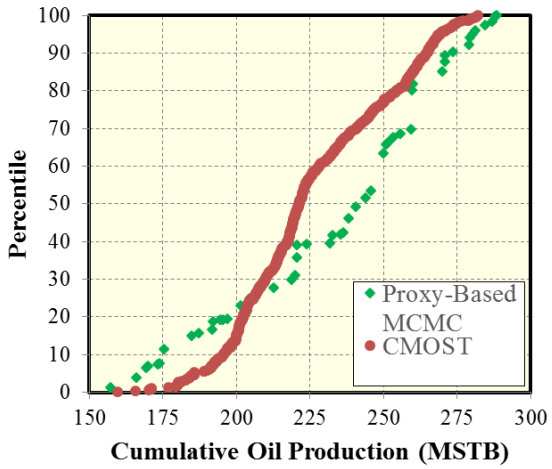
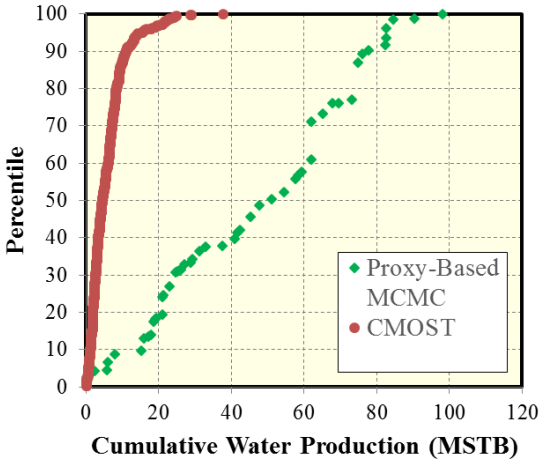


Figure 4.21: Permeability solution from the proxy-based MCMC workflow and CMOST.

The inaccurate PPD of history-matching solutions will bias the probabilistic forecasts and likely result in the wrong decisions in reservoir management. Figure 4.22 displays the forecasted s-curves by proxy-based MCMC workflow and CMOST. We can see that the PPD by CMOST that misses some of the solution leads to the overconfidence in the oil production uncertainty, the optimistic forecasting of low water production, and the optimistic forecasting of high oil recover factor, as shown in Figure 4.22(a), (b), and (c), respectively.

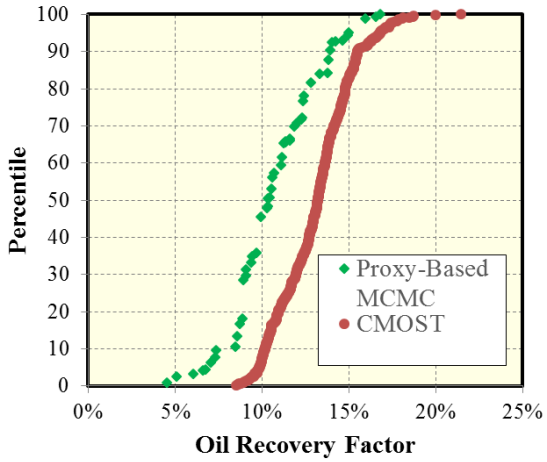


a) S-curve of cumulative oil production after 8000 days



b) S-curve of cumulative water production after 8000 days

Figure 4.22: continued next page.



(c) S-curve of oil recover factor after 8000 days

Figure 4.22: Comparison of the forecasted S-curves between the proxy-based MCMC and CMOST

4.5 Comparison of The Workflow with The Direct MCMC

To compare the performance of the proxy-based MCMC method, the direct MCMC method is implemented to history match the BHP profile of the same case study by varying the five significant parameters. The maximum limit of simulations runs is set to be 10000 simulation runs. It is found that the direct MCMC can merely find three unique solutions. The comparison of the efficiency between the proxy-based and the direct MCMC methods are provided in Table 4.4. As shown, the computational cost, in terms of total simulation runs per a solution found, of the direct MCMC method is more than 37 times higher than that of the proxy-based MCMC.

Table 4.4: Comparison of the performances between the proxy-based MCMC workflow and the direct MCMC method

Parameter	Proxy-Based MCMC workflow	Direct MCMC method
Simulation runs to improve proxy	2675	-
Simulation runs to explore final proxy	2000	-
Total simulation runs	4675	10000
History matching solutions	52	3
Total simulation runs/solutions	89.9	3333.3

In addition, we validate the convergence of the MCMC by monitoring the coefficient of variation (CV) of all the significant parameters. The CV is a dimensionless statistical parameter that is suitable to measure variation for the data set with different units or widely different means. The following equation is used to calculate CV (Everitt, 1998)

$$C_v = \frac{\sigma}{\mu}, \tag{4.6}$$

where C_v is the coefficient of variation, σ is the standard deviation, and μ is the mean value.

The CV vs iteration plots are shown in Figure 4.23. The majority of the plots seem to establish downward trends at the late part of chains, particularly in Fig. 17(d). The changing CVs suggest that the direct MCMC chain has not yet converged even after 10000 iteration. The high number of iterations required supports the previous observation, i.e., the prior knowledge is highly uncertain and high computational cost for history matching is expected.

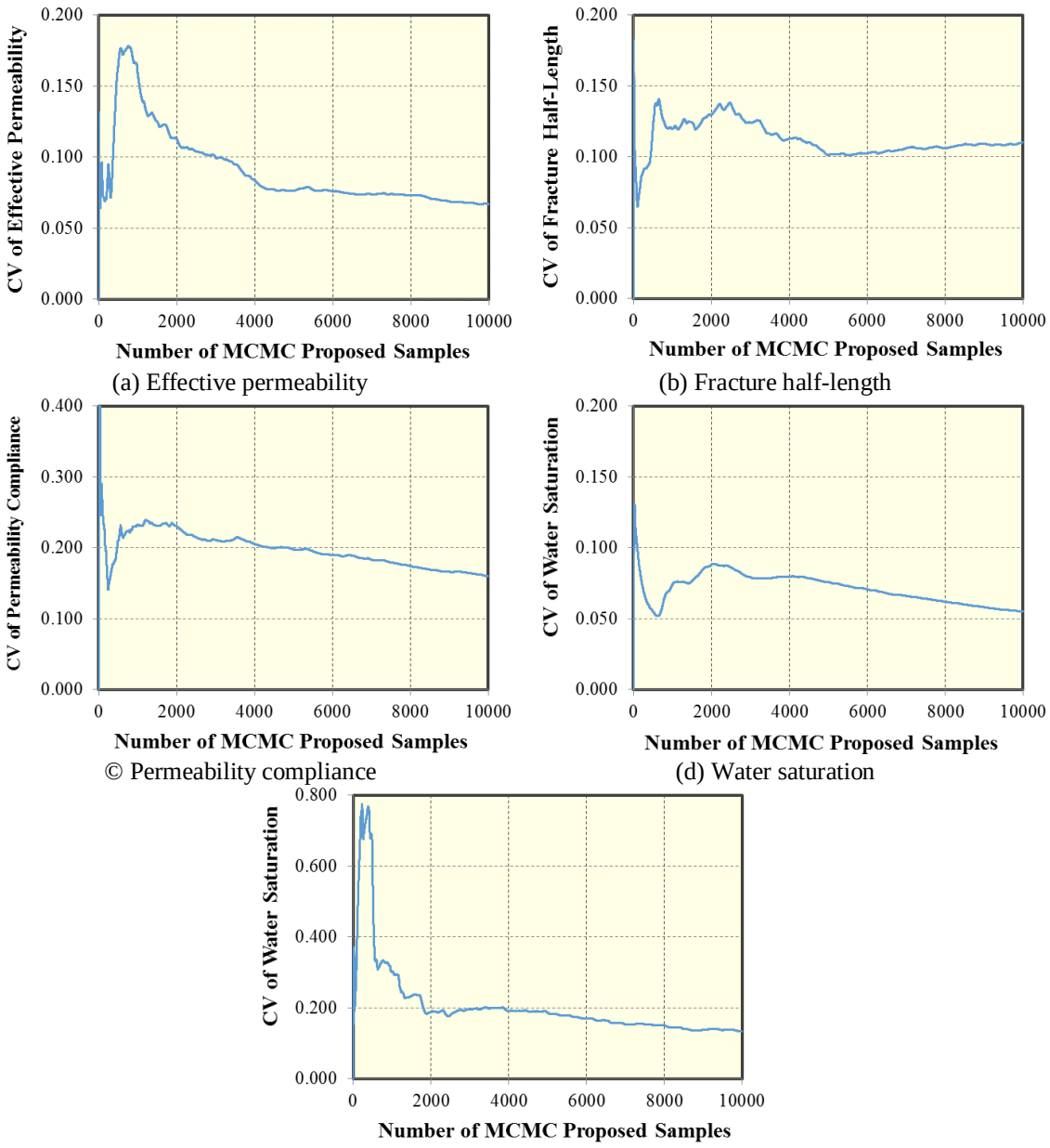


Figure 4.23: Coefficient of variance for the four uncertain parameters against the proposed MCMC samples

4.6 CONCLUSIONS

We have applied the proxy-based and the direct MCMC workflows to history match BHP and water production data from a shale oil well in Vaca Muerta formation. The key findings are summarized as follows:

1. The proxy-based MCMC workflow can history match and construct the approximate PPD for the shale oil case in 100 iterations for a total of 4675 simulation runs. Thus, the proxy-based MCMC workflow has been proved to be effective for the history matching of unconventional reservoirs. However, the accuracy of the final proxy does not reach the satisfactory level. For the future work, the improvement of the accuracy may be done via using other types of proxy such as kriging or redefining the prior uncertainty ranges
2. The efficiency, in term of solutions found per total simulations run, of CMOST is substantially higher than the proxy-based MCMC. However, in terms of accuracy, the former misses several history-matching solutions that are found by the latter.
3. The direct MCMC workflow finds only few solutions thus cannot construct the PPD despite 10000 simulation runs are performed. Hence, the direct MCMC workflow is not recommended for history matching due to the enormous computational cost and it is preferable to use the proxy-based MCMC workflow.
4. According to the history matching, the SRV is found to be on the low side of the expectation due to the low values of the history-matched fracture half-length and porosity.
5. With the producing BHP of 500 psi, the P10-50-90 oil EUR of the Vaca Muerta well is forecasted to be 175-244-274 MSTB and the P10-50-90 oil recover factor of the well is forecasted to be 8%-10%-14%.

Chapter 5: Integration of the AHM Workflow, Microseismic, and EDFM to History Match a Shale Oil Field Case

In this chapter, we demonstrate the synergy of EDFM and the proxy-based MCMC workflow to history match a shale play with complex fracture. The field case is the same shale oil well as in the previous chapter. However, the microseismic data are now available and accounted to constrain the uncertain geometries of hydraulic fractures. We use Design of Experiment (DoE) as the screening tool to reduce the number of uncertain parameters to the manageable level. Then, the integrated workflow is applied to history match this complex fractured reservoir. The implementation is efficient and successful. In addition to reducing the uncertainties of reservoir properties, the knowledge of fracture network is more certain after history matching.

5.1 Field Case Study and Microseismic Data in Shale Oil Reservoir

The field case is the same as that in section 4.1 except that the additional microseismic data is available and is incorporated in fracture modelling. The well was planned to complete with nine fracturing stages but two of them experienced execution failures, thus requiring re-fracturing. The fracturing fluid was comprised of a combination slick water, linear gel and crosslinked gel fluid carrying 50/150, 30/50, 30/60 proppants respectively.

Microseismic data was acquired during the completion of each fracturing stage. Figure 5.1 shows the map view of the treatment well, the monitor well, and the microseismic events for all monitored stages. According to visual inspection of the microseismic distribution, fracture geometry of each stage is varied along the lateral section. Stage 1 and 9 indicate planar geometry while stage 2, 3, 4, and 5 exhibit multiple

or complex fracture networks. In addition, stage 2, 3, and 4 possibly make contact with the monitoring well. Relatively few microseismic events are observed in stage 6, 7, and 8.

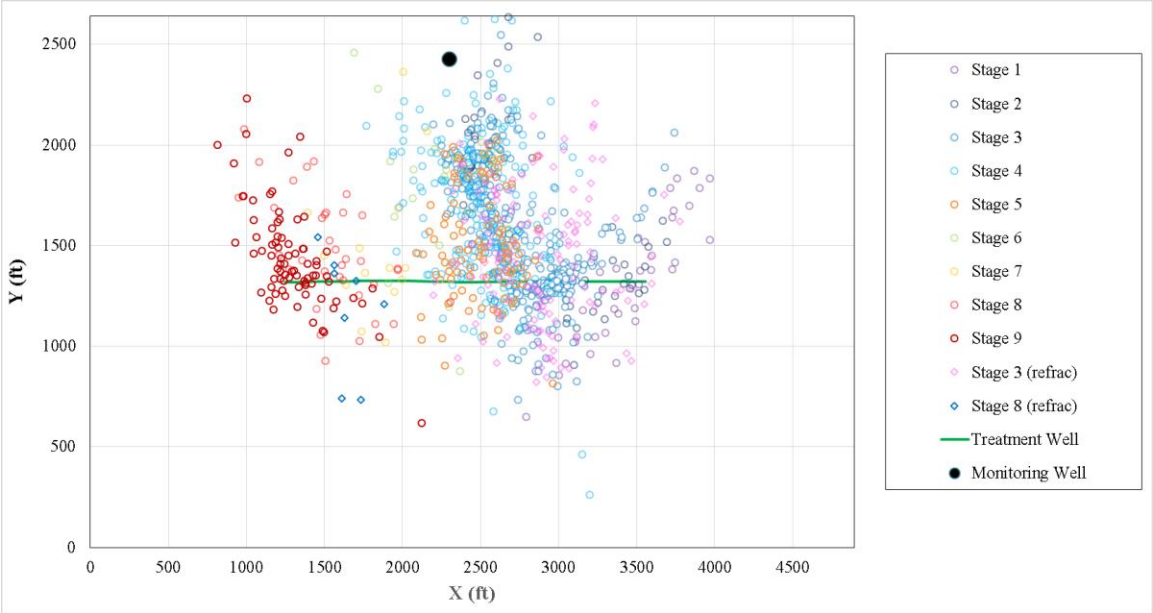


Figure 5.1: Treatment well, monitoring well, and microseismic events.

Figure 5.2 displays the 28 interpreted discrete fractures. 27 of them are hydraulic fractures, displayed as black lines, and the rest is a pre-existing fracture, shown as the red line. These fractures are modelled while honoring the azimuth and the density of microseismic events. The maximum possible lengths of the hydraulic fractures are shown as dash lines whereas their effective lengths, which are uncertain, are shown as solid lines The effective length of every hydraulic fracture in this case study is formulated by multiplying its corresponding maximum length with a scaling factor. This fracture scaling factor will be treated as an uncertain parameter in the subsequent history matching process. The effective length of the pre-existing fracture is however assumed to be constant and the conductivity of this fracture is assumed to be lower than those of the hydraulic ones.

The basic reservoir and fracture parameters are given in Table 5.1. It should be noted that the permeability compliance is assumed to be 0 in this study because the current version of the EDFM preprocessor cannot handle the dynamic permeability. To compensate the lack of this parameter, the prior uncertainty of permeability is assumed to have a broad range, from 100 nD to 0.1 mD. The eight uncertain parameters as well as their prior distributions are summarized in Table 5.2.

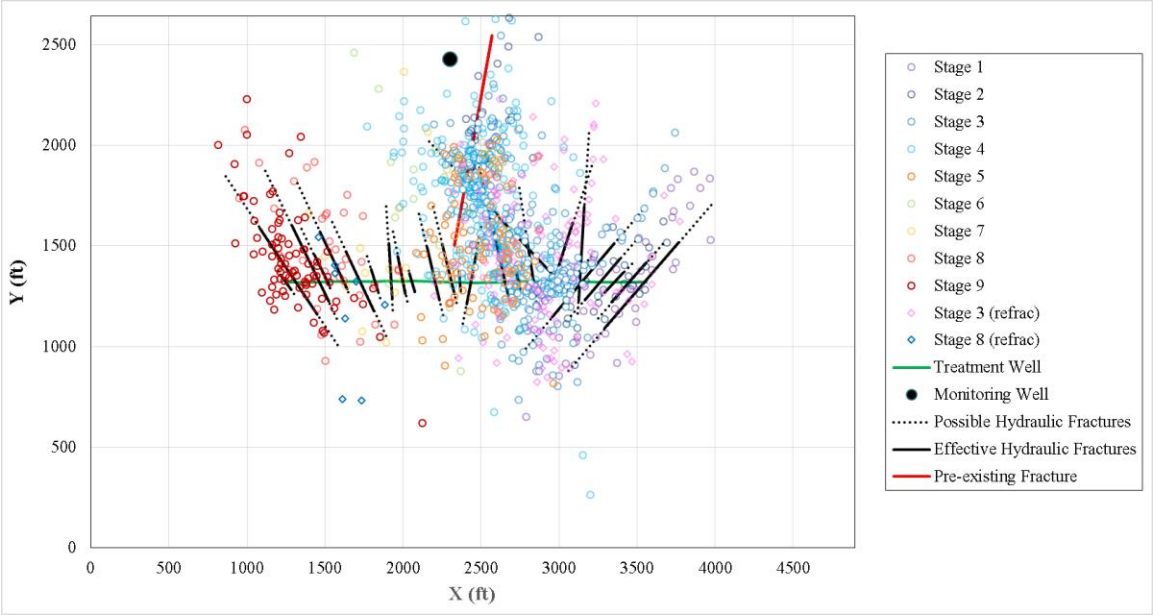


Figure 5.2: Interpreted fracture geometry based on the microseismic data.

Table 5.1: Summary of given reservoir properties

Parameter	Unit	Value
Reservoir pressure	psi	9500
Reservoir datum depth	ft	10000
Reservoir length	ft	4897
Reservoir width	ft	2640
Permeability Compliance	1/psi	0

Table 5.2: Summary of the eight uncertain parameters and their prior distributions

Code	Uncertain parameter	Unit	Distribution	Min	Max
A	Permeability	mD	Uniform	0.0001	0.1
B	Water Saturation	-	Uniform	0.2	0.5
C	Porosity	-	Uniform	0.001	0.08
D	Total Compressibility	1/psi	Uniform	5.8E-06	9.0E-06
E	Hydraulic Fracture Conductivity	md-ft	Uniform	100	1000
F	Thickness	ft	Uniform	275	315
G	Pre-existing Fracture Conductivity	md-ft	Uniform	0.1	10
H	Fracture Scaling Factor	-	Uniform	0.1	1

The reservoir model is built using a commercial reservoir simulator (CMG-IMEX, 2014). To efficiently model this complex fracture geometry into the reservoir model, EDFM, as opposed to LGR, is implemented. The more explanation of EDFM can be found in the next section.

5.2 EMBEDDED DISCRETE FRACTURE MODEL (EDFM)

EDFM is an efficient method to model complex fracture in a reservoir simulation. In this method, each fracture plane is discretized by matrix cell boundaries and its influence is explicitly modeled through the transmissibility factor calculation. EDFM automatically generates virtual grids and assigns non-neighboring connections (NNCs) for these cells to account for the transmissibility factors associated with fractures including the flows between matrix-fracture, fracture segment–fracture segment, and well-fracture. Insertion of fractures is intrusive for a reservoir model and does not require local grid refinement (LGR); hence, using EDFM to model fractures is more efficient than the unstructured grid technique. On the other hand, explicit calculation of fracture influence in EDFM provides more accurate fracture model than the traditional continuum approach such as dual-porosity or dual-permeability model.

The more detailed calculation of NNC transmissibility factors are available in Xu et al. (2017). In brief, the calculation is mainly based on two-point flux approximation where the flow is governed by Eq. 5.1. The EDFM only calculates the phase-independent part of the transmissibility while the relative mobility is calculated by the reservoir simulator.

$$q_l = \lambda_l T_{NNC} \Delta P , \tag{5.1}$$

where q_l is the volume flow rate of phase l, λ_l is the relative mobility of phase , T_{NNC} is the NNC transmissibility factor, and ΔP is the potential difference between two points.

For the matrix-fracture connection, the NNC transmissibility calculation can be generally expressed by Eq. 5.2

$$T_{NNC} = \frac{k_{NNC} A_{NNC}}{d_{NNC}} , \tag{5.2}$$

where k_{NNC} is the matrix permeability in the direction normal to the fracture plane, A_{NNC} is the contact area of the fracture plane inside the matrix block, and d_{NNC} is the average normal distance from matrix block to fracture plane.

For the fracture segment- fracture segment connection, T_{NNC} is the harmonic average between the transmissibility factors of the two segments. The formulation for calculating transmissibility of the fracture segment also depends on whether the two segments belong to an individual fracture or belong to two different fractures. The more detailed explanation can be found in Xu et al. (2017).

For the well-fracture intersection, the effective well index is calculated and assigned to the intersection. The formulation used for the well index calculation is Peaceman (1983) equation, as shown in Eq. 5.2

For the well-fracture intersection, the effective well index is calculated and assigned to the intersection. The formulation used for the well index calculation is Peaceman (1983) equation, as shown in Eq. 5.2

$$WI_f = \frac{2\pi k_f w_f}{\ln(\frac{r_e}{r_w})}, r_e = 0.14\sqrt{L_S^2 + H_S^2}\,, \tag{5.3}$$

where k_f is the fracture permeability, w_f is the fracture aperture, L_s is the length of the fracture segment, H_s is the height of the fracture segment, r_e is the effective radius, and r_w is the well-bore radius.

To ensure that EDFM reproduces the same simulated result as LGR, a validation test is performed using a simpler, bi-wing case. The validating case is analogous with the case in Chapter 4. Its basic reservoir properties are provided in Table 5.3. For the validation purpose, the simulated BHP by EDFM will be compared against that by LGR where an oil rate profile is used as the input entry. Figure 5.3 displays the simulated BHP of the two methods. As can be seen, their BHP profiles almost perfectly align with each other. Moreover, the simulated reservoir pressure map by LGR in Figure 5.4 is reasonably consistent with that be EDFM in Figure 5.5. Therefore, EDFM shows excellent agreement with LGR and will be used to model the fracture for the rest of the study.

Table 5.3: Summary of given reservoir and fracture parameters of the validating case

Parameter	Unit	Value
Reservoir pressure	psi	9500
Reservoir datum depth	ft	10000
Reservoir length	ft	4897
Reservoir width	ft	2640
Number of fracture	-	53
Fracture spacing	ft	59
Permeability	mD	0.0867
Thickness	ft	295
Fracture Half-Length	ft	137
Permeability Compliance	1/psi	0
Water Saturation	-	0.409
Porosity	-	0.0045
Formation Compressibility	1/psi	1.00E-6
Fracture Conductivity	md-ft	500

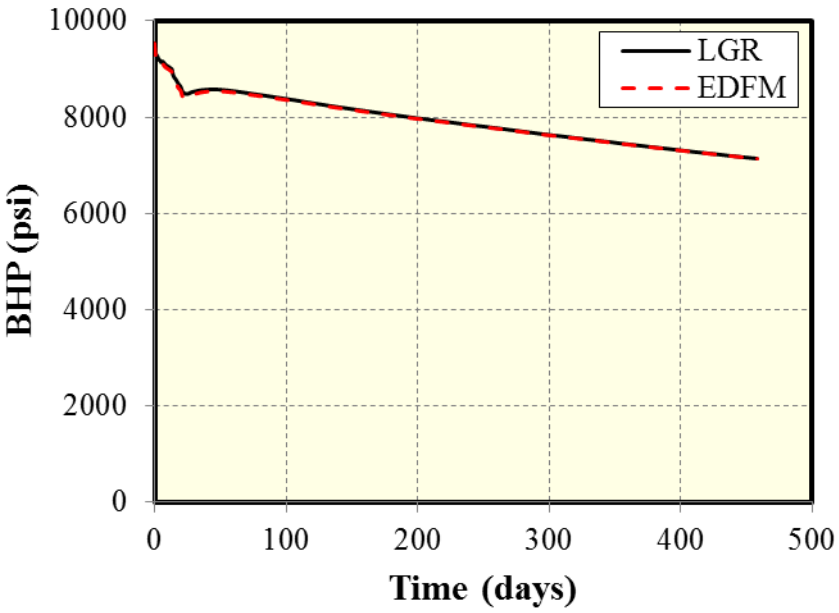


Figure 5.3: Comparison of the simulated BHP profiles by LGR and EDFM

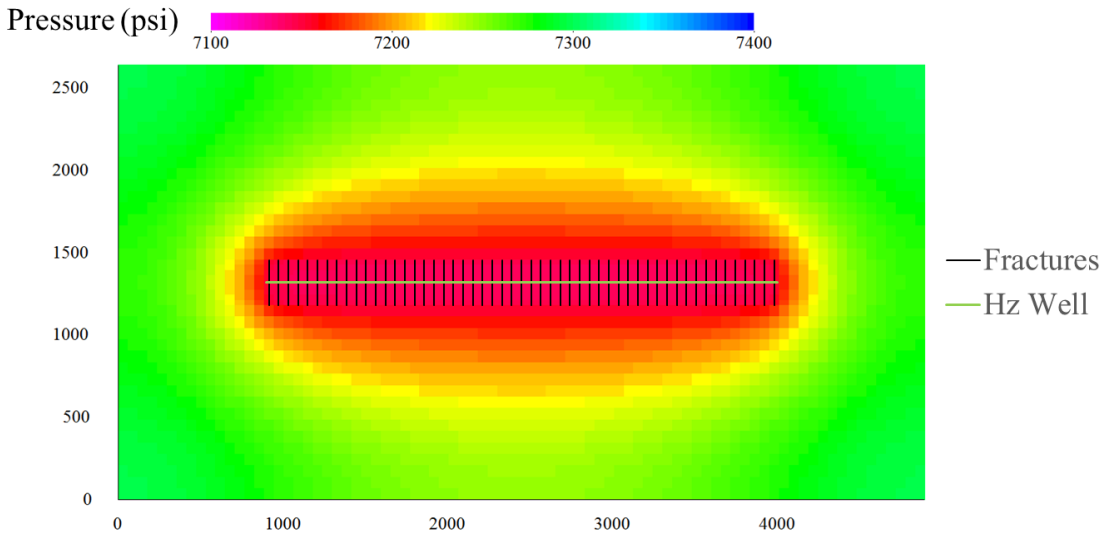


Figure 5.4: Reservoir pressure map after 458 days of the validating case using LGR

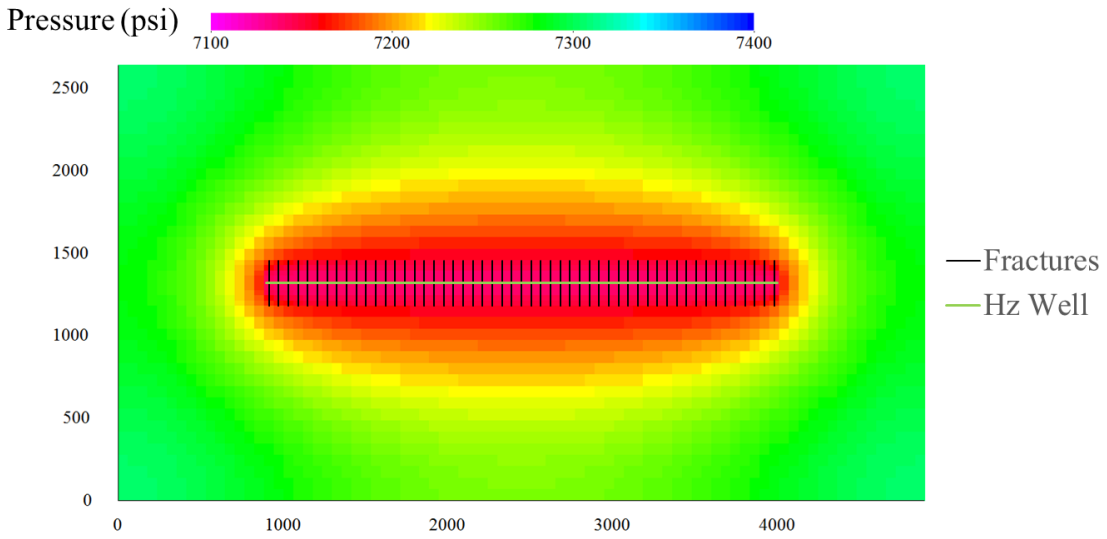


Figure 5.5: Reservoir pressure map after 458 days of the validating case using EDFM

5.3 METHODOLOGY

The methodology is summarized as the flow chart in Figure 5.6. The workflow is essentially similar to Figure 4.3 and the explanation is referred to Section 4.2. The exceptions are as follows. Firstly, the EDFM preprocessor is required, in addition to the reservoir simulator, for generating simulated results. Secondly, the limit of maximum iteration number is reduced to be 20. Thirdly, the number of solution points sampled from the final proxy by MCMC is reduced to 1000. The last two changes are made to reduce the computational time since the complex-fracture case in this study is more computationally expensive than the bi-wing case in Figure 4.3. Lastly, due to the lack of permeability compliance as an uncertain parameter, highly accurate history matching is not practically possible to achieve. Therefore, we compromise the quality of history matching solutions by using less than 400 psi BHP RMSE and less than 40 STB/day water RMSE as the cut-offs.

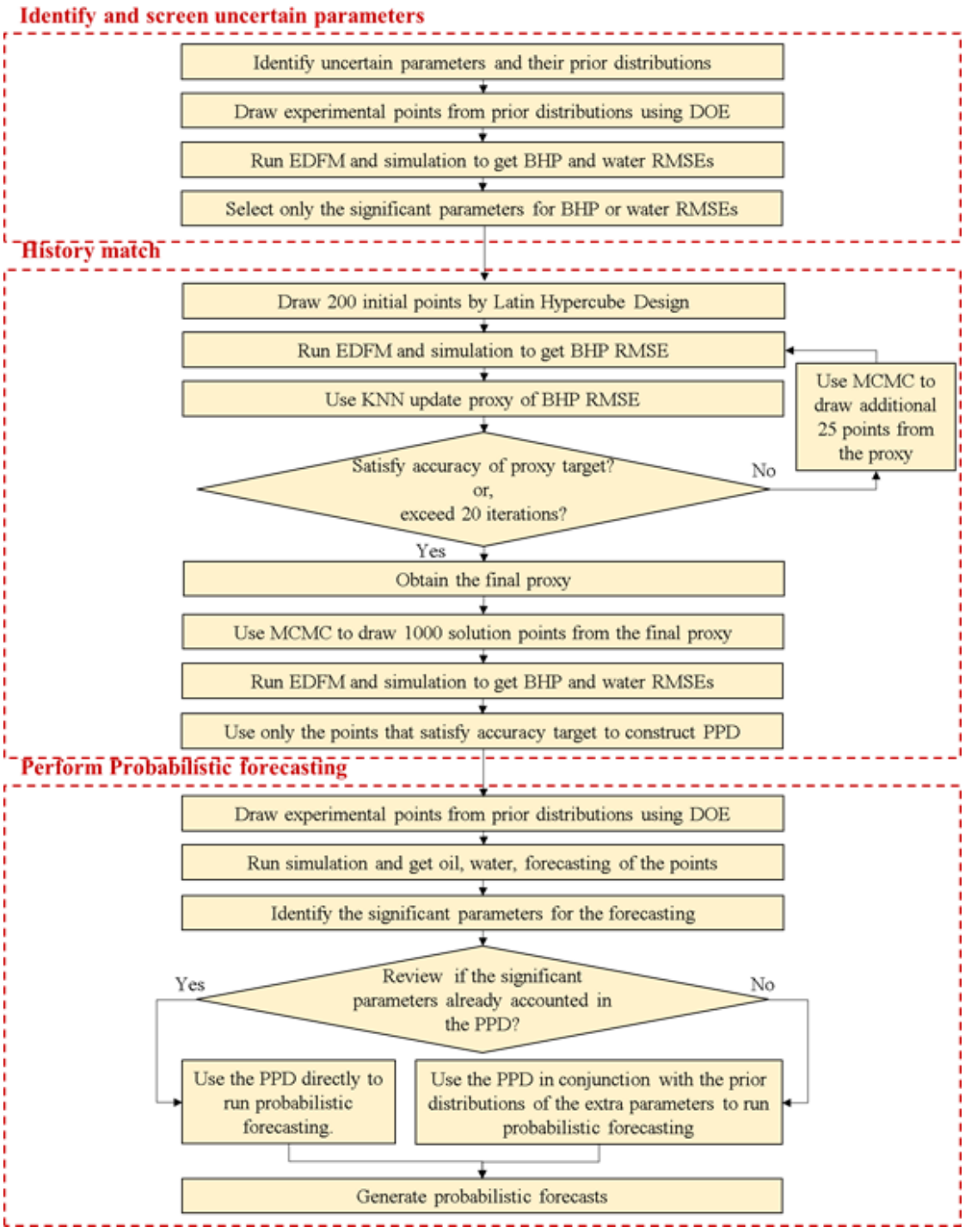


Figure 5.6: Flowchart of the AHM workflow using proxy-based MCMC applied on the shale oil with complex fracture field case.

5.4 RESULTS

During DOE step to screen significant parameters, we use a two-level factorial design to generate 128 reservoir model realizations for the eight parameters in Table 5.2. The influences of parameters on the BHP RMSE are shown the Pareto charts in Figure 5.7. The definition of each term on the x-axe is similar to that in Table 5.2. As can be seen, BHP RMSE is mostly impacted by the terms AC, C, ABCH, BCH, and A. Figure 5.8 displays the Pareto charts of the water RMSE in the similar fashion. We can see that water RMSE is highly influenced by the terms B, AH, H, ABH, and BH. Therefore, the A, B, C, and H parameters (permeability, water saturation, porosity, and fracture scaling factor) are the most significant parameters responsible for history matching quality of the BHP and water profiles while the other four parameters are kept constant at reference values in the history matching step.

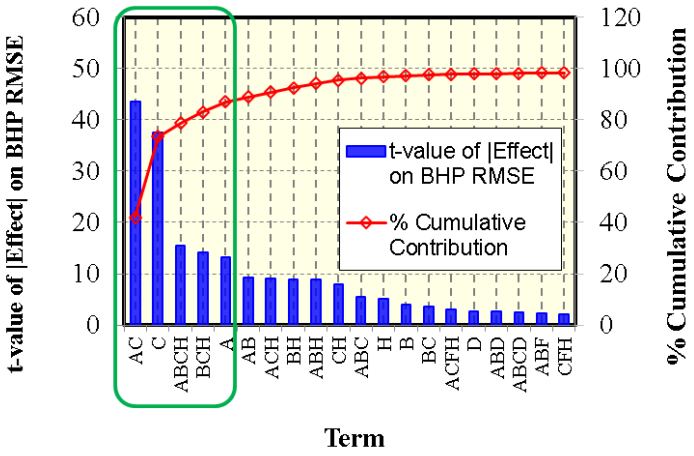


Figure 5.7: Pareto charts of BHP RMSE according to the two-level factorial design.

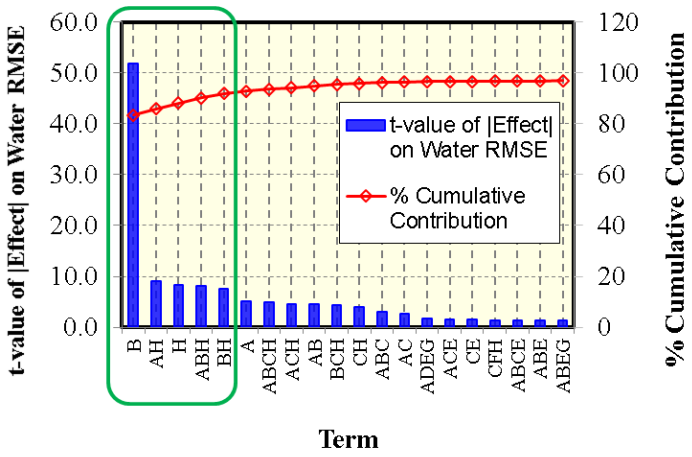
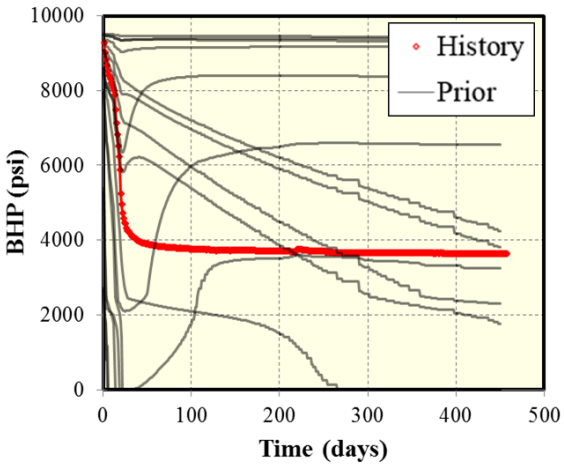
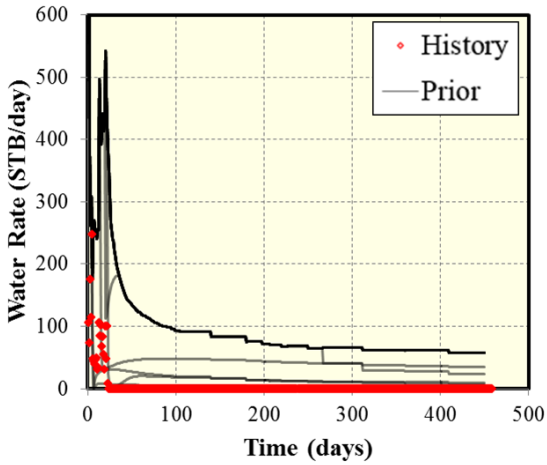


Figure 5.8: Pareto charts of water RMSE according to the two-level factorial design.

To display the diversity of the four significant parameters, the forecasts based on a full factorial design for the total of 16 cases are displayed in Figure 5.9. As the plots illustrate, the simulated BHP profiles in Figure 5.9(a) are greatly diverse, i.e., the pressure profile remains nearly flat in a case while the pressure profile of another case rapidly drops to zero. The vast difference of the BHP responses indicates the high uncertainty in the prior distributions so high computational cost is probably expected for history matching.



(a) Historical and simulated bottomhole pressure profiles



(b) Historical and simulated water rate profiles

Figure 5.9: Historical profiles and the possible 64 simulated profiles based on the prior distribution of the five significant parameters using full factorial design

After the limit of 20 iterations is exceeded, the final BHP RMSE proxy is attained. We then use MCMC to sample 1000 points that have lowest RMSEs from the final proxy. Out of these samples, 53 realizations are found to be the actual solutions, (less than 400 psi BHP RMSE and less than 40 STB/day water RMSE) after confirmation by running reservoir simulation. The final solutions are summarized as the parallel coordinate plot shown in Figure 5.10. Each of the green lines represents one of the history-matched realizations connecting the four parameters in the first four vertical axes and the sampling frequency in the fifth axis.

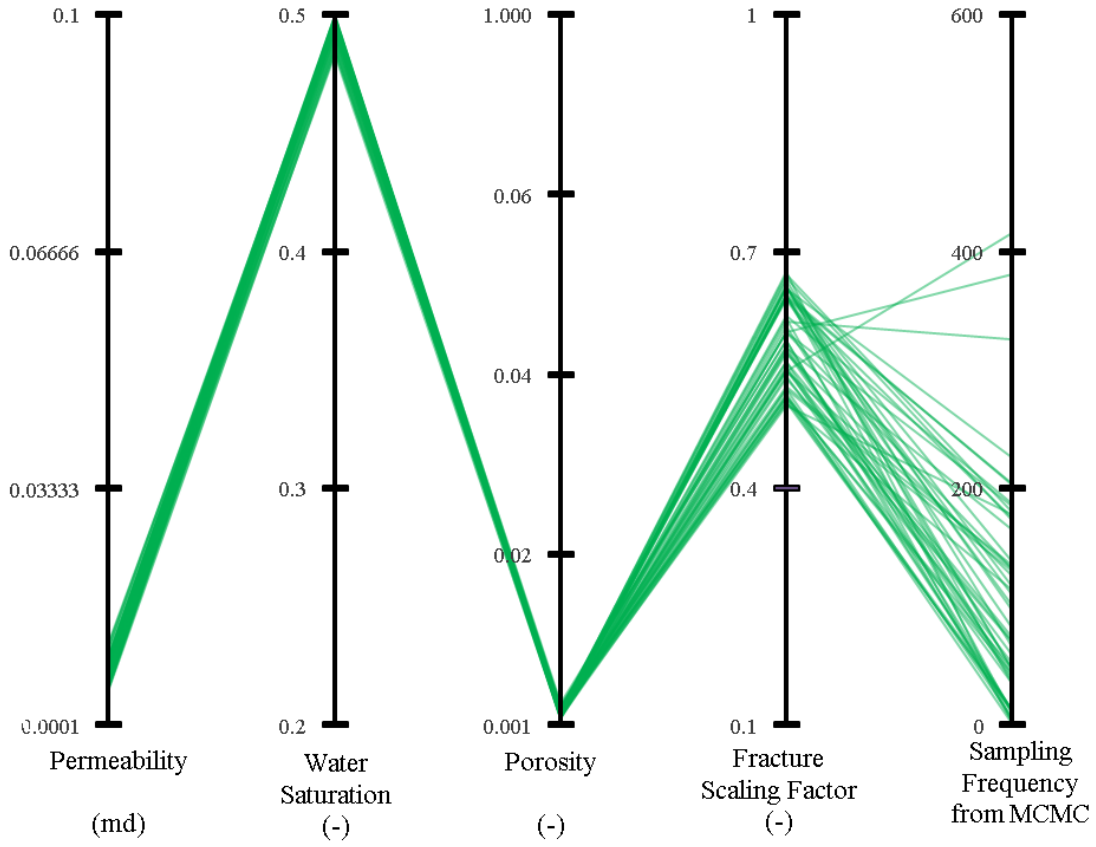
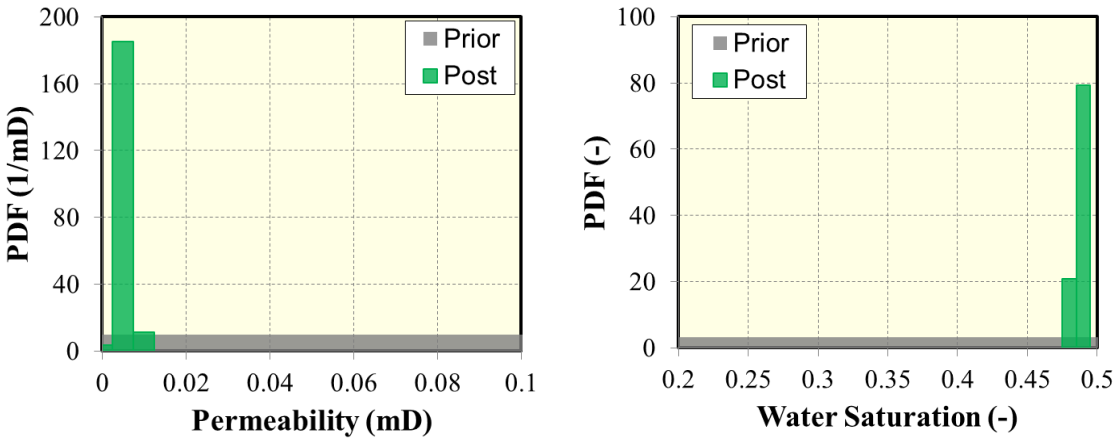


Figure 5.10: Parallel coordinate plot of 53 history-matched solutions found by the proxy-based workflow.

These 53 solutions are projected onto each of the parameters and plotted as the histograms displayed in Figure 5.11. As can be observed, history-matched permeability, porosity, and water saturation ranges have very narrow spreads which indicates the difficulty to search solutions. According to the history-matched fracture scaling factor range, the effective fracture lengths are expected to range from 50% to 65% of the possible fracture lengths. Comparing Figure 5.11(b) with Figure 4.10(d), the water saturation of this case is relatively higher than that of the bi-wing case. The reason for this discrepancy is partially attributed to the negligence of the permeability compliance. The BHP history

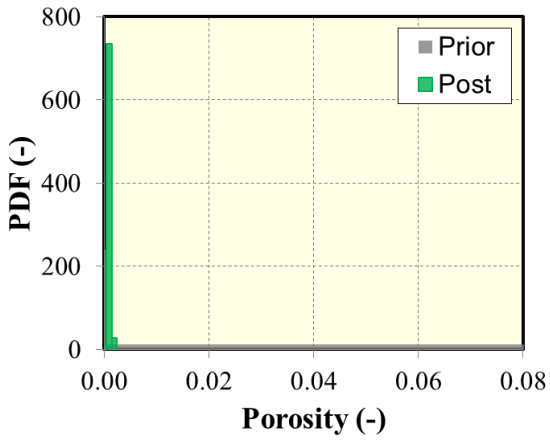
matching for this field case ideally should have accounted for the oil permeability reduction as a function of pressure depletion. Without the option of the dynamic permeability, the water saturation is forced to increase for reducing the oil permeability. Moreover, the lack of this key uncertainty makes it impossible to find a solution that simultaneously satisfy both of the BHP and water RMSE cut-offs. In other words, high quality of BHP history matching must be attained at the expense of lower quality of water history matching. The comparison between the prior and posterior simulated results is displayed in Figure 5.12. As shown, the history-matched BHP profiles have reasonably good agreement with the BHP history whereas the posterior water profiles are deviated from the historical water profiles.



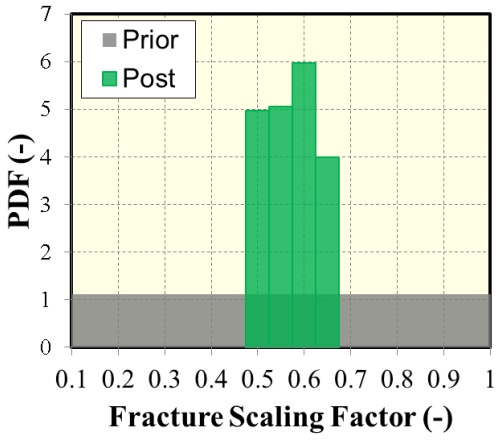
(a) Permeability

(b) Water saturation

Figure 5.11: continued next page.

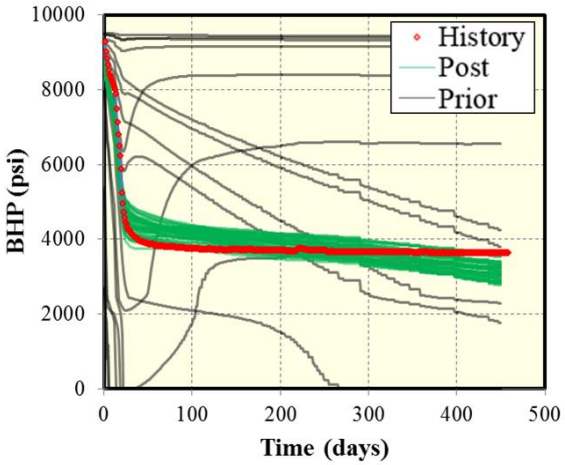


(c) Porosity

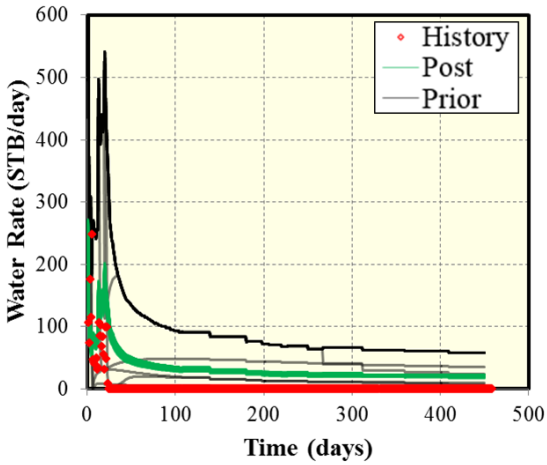


(d) Fracture Scaling Factor

Figure 5.11: Comparison between prior and posterior probability function of the four significant uncertain parameters using the proxy-based MCMC



(c) Historical and simulated bottomhole pressure profiles



(d) Historical and simulated water rate profiles

Figure 5.12: Comparison between prior and posterior simulated results

Figure 5.13 displays the reservoir pressure map after 458 days of a history-matched realization that has the fracture scaling factor of 51%. The majority of pressure drop occurs near the effective hydraulic fractures. While microseismic data can estimate maximum possible fracture lengths, effective fracture lengths can only be determined by history

matching process. The more information of the effective fracture geometry will significantly contribute to reservoir management and well spacing strategy.

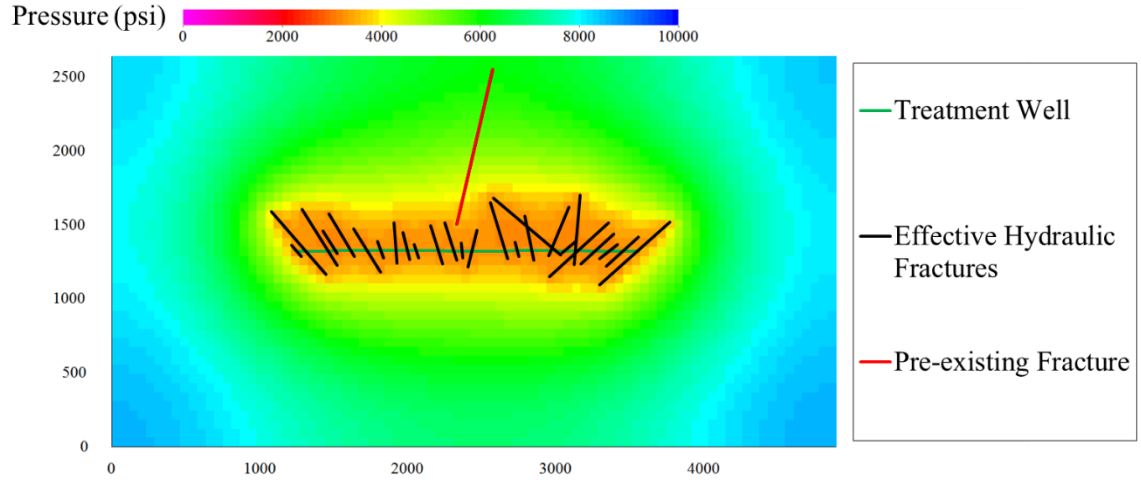


Figure 5.13: Reservoir pressure map of the complex fracture case after 458 days

For the probabilistic forecasts, the well is assumed to produce with BHP of 500 psi until 8000 days. We use two-level factorial design to analyze the influences of all the eight parameters in Table 5.2. The Pareto charts for the cumulative oil production, cumulative water production, and oil recovery factor forecasts are presented in Figure 5.14, Figure 5.15, and Figure 5.16, respectively. As highlighted by the green rectangles, the most influential parameters for the three forecasts are A, B, and C. Since the set of these parameters is a subset of the parameters influential for history matching quality, the final PPD constructed in the history matching process can be directly used for the probabilistic forecasting step

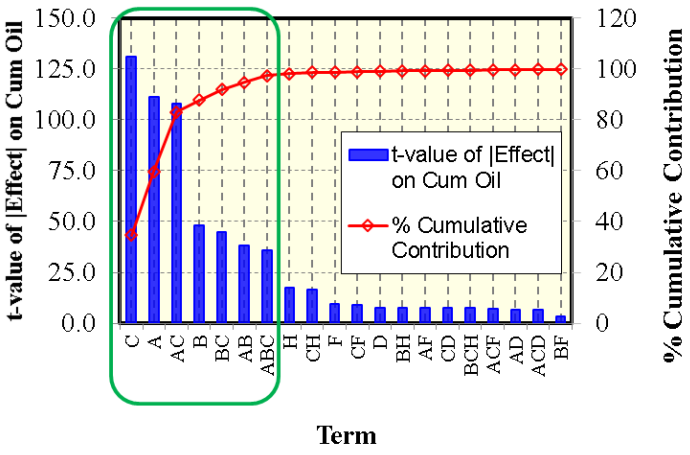


Figure 5.14: Pareto charts of cumulative oil forecast according to the two-level factorial design.

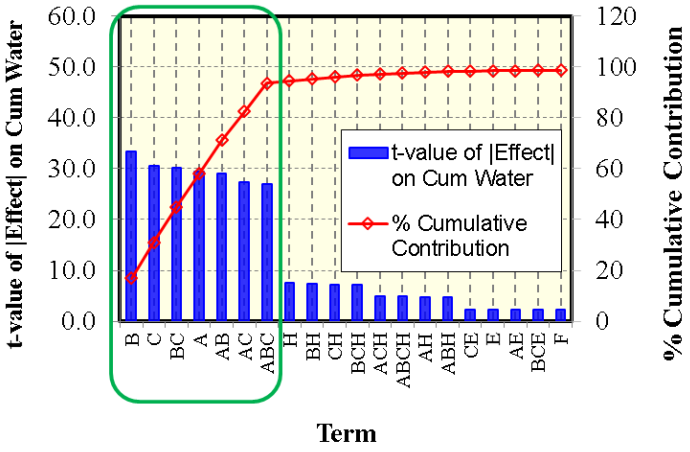


Figure 5.15: Pareto charts of cumulative water forecast according to the two-level factorial design.

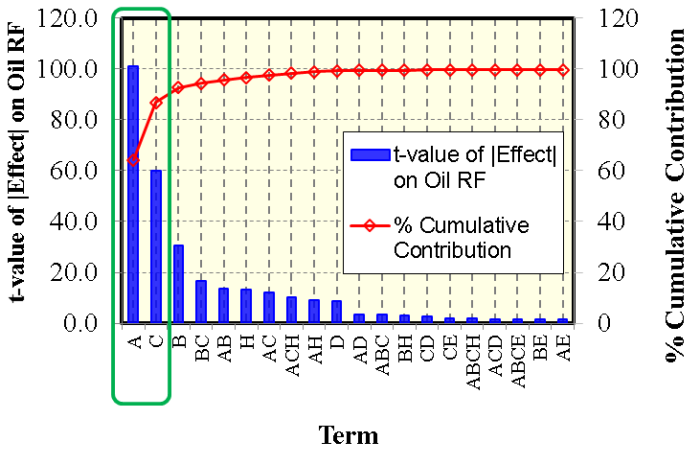
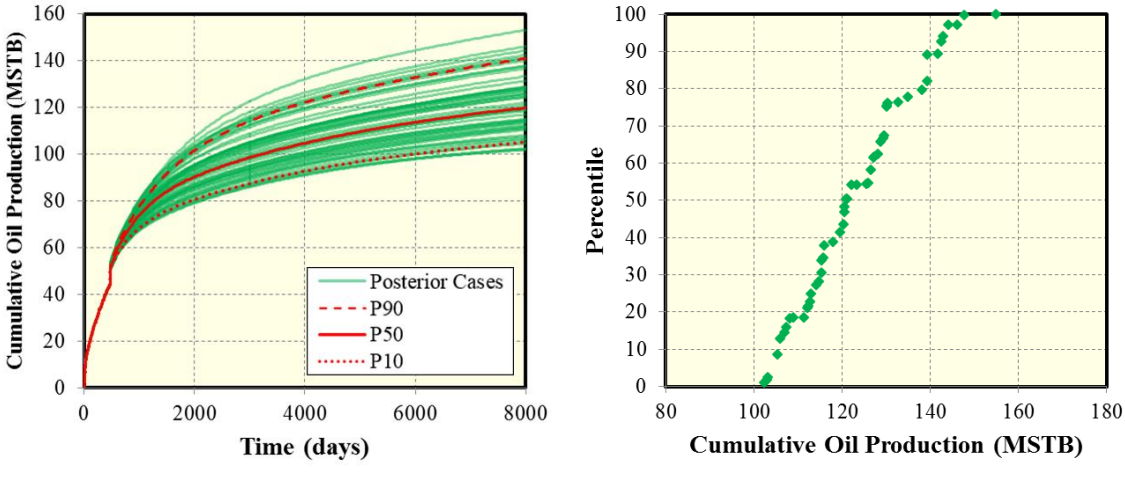


Figure 5.16: Pareto charts of oil recovery forecast according to the two-level factorial design.

Figure 5.17, Figure 5.18, and Figure 5.19 display the posterior forecasts of cumulative oil production, cumulative gas production, and oil recover factor, respectively. The probabilistic forecasts versus time are plotted in the left panels and the s-curves of the three forecasted values after 8000 days are displayed in the right panels. As shown in Figure 5.17, the P10-50-90 cumulative oil production is 106-121-143 MSTB. The ratio of P90/P10 of 1.3 indicates moderate uncertainty in the oil production forecast.

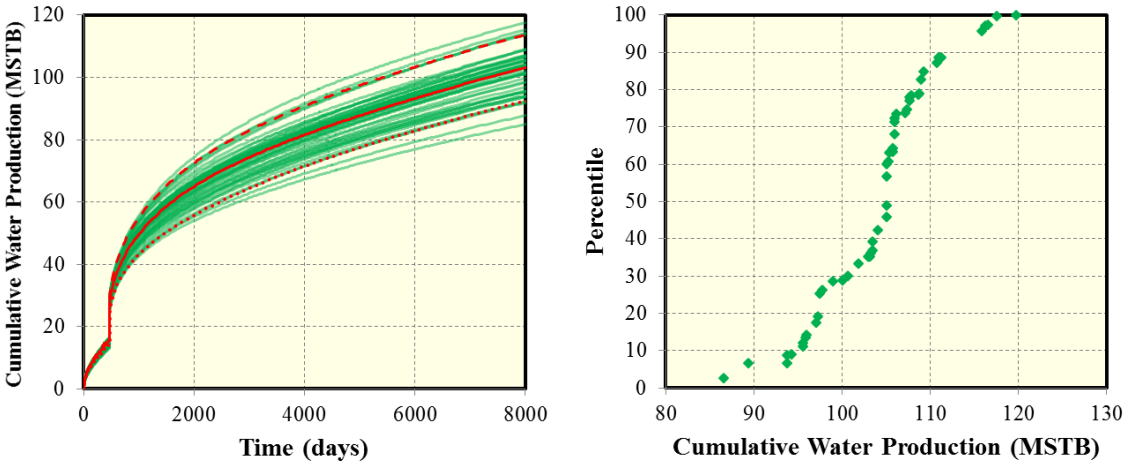
The water production is forecasted to be between 94 and 106 MSTB, as presented in Figure 5.18. The forecast, however, is suspiciously overestimated since the history-matched water profiles in Figure 5.12 are inaccurately higher than the history. The P10-50-90 oil recover factor of the Vaca Muerta well is 22%-25%-28%, as shown in Figure 5.19. However, these uncertainty quantifications are supposed to change if accounting the permeability compliance as an uncertain parameter. We recommend revisiting and incorporating the pressured dependent permeability in this case once the new version of EDFM preprocessor that is capable of handling the dynamic permeability is released.



a) Cumulative oil production vs time

b) S-curve of cumulative oil production after 8000 days

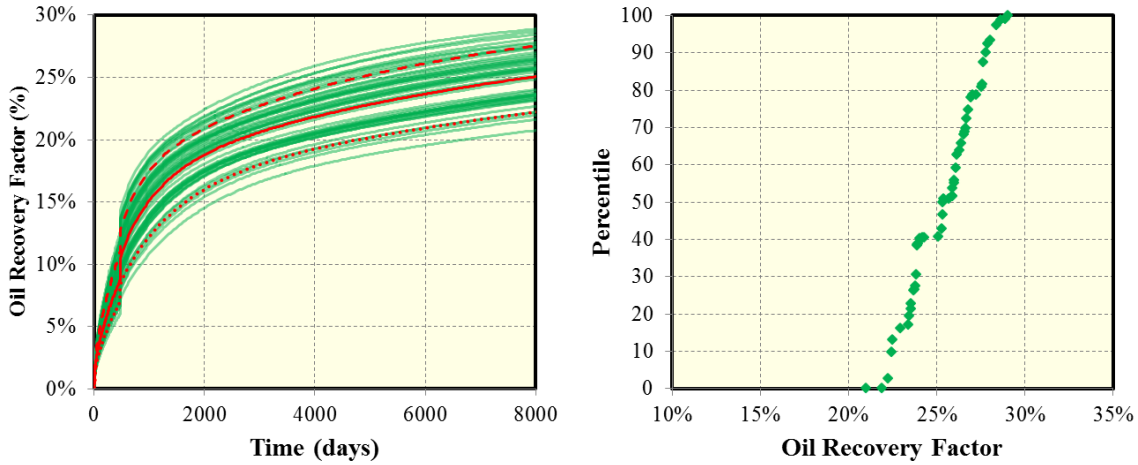
Figure 5.17: Posterior cumulative oil production forecasts.



a) Cumulative water production vs time

b) S-curve of cumulative water production after 8000 days

Figure 5.18: Posterior cumulative water production forecasts.



a) Oil recovery factor vs time

b) S-curve of oil recovery factor after 8000 days

Figure 5.19: Posterior oil recover factor forecasts.

5.5 CONCLUSIONS

We have applied the workflow using proxy-based MCMC to history match BHP and water production data from a shale oil well in Vaca Muerta formation. The EDFM preprocessor is also integrated in the workflow to efficiently model the complex fracture geometry for the reservoir model. The findings are summarized as follows:

1. The proxy-based MCMC and the EDFM preprocessor is successfully synergized using MATLAB and VBA. This integrated workflow combines the efficiency to history match of the former and the efficiency to model discrete fractures of the latter, thus providing the capability of history matching reservoir with complex fracture geometry.
2. The integrated workflow can determine the likelihood of complex fracture geometry. In this case study, a single fracture scaling is used to govern the generation of fracture geometry realizations. Determining the likelihood of each fracture scenario conditioned to the production data is particularly useful to quantify uncertainty and develop optimization plan.

3. According to the given assumptions, the integrated workflow can history match and construct the approximate PPD for the shale oil case in 20 iterations for a total of 2675 simulation runs. The probability forecast, after history matching, indicates that the P10-50-90 oil EUR of the Vaca Muerta well is 106-121-143 MSTB and the P10-50-90 oil recover factor of the well is forecasted to be 22%-25%-28%.

Chapter 6: Summary Conclusions, and Recommendations for Future Work

6.1 Summary and Conclusions

The investigation of AHM workflow using the proxy-based MCMC is performed on the synthetic simple case. Several insights are discovered for this black-box process and these findings are incorporated to improve the performance of the workflow. The EDFM preprocessor is also integrated in the proxy-based MCMC workflow to efficiently model several scenarios of the complex fracture geometry during history matching. The integrated workflow is used to history match the oil shale field case in Vaca Muerta Formation, Argentina where the microseismic data are used to constrain the uncertain geometries of fractures process. The history-matching using the workflow is efficient and successful. In addition, the comparative study between the workflow and CMOST software is conducted to compare the accuracy and efficiency. The key findings are concluded as follows:

- The comparative study of proxy type indicates that, the most accurate proxy is kriging, follow by KNN and cubic proxies. The least accurate is quadratic proxy. On the other hand, in terms of computational time, quadratic proxy is the cheapest, cubic proxy is the second, KNN proxy is the third, and kriging proxy is the fourth. In our experience, KNN proxy is an acceptable proxy providing slightly less accuracy than kriging proxy but requires substantially less computational time.
- It is also found that measurement errors smear the precision of the RMSE objective function. In turn, the lower precision reduces the urgency of using very accurate proxy like kriging. In some cases, high measurement errors might even allow a rough model like quadratic to adequately describe the RMSE response surfaces.

- According to the study in the synthetic simple case, R^2, between the proxy responses of the current iteration and the actual responses of the following iteration is recommended to use as the stop criterion of the proxy-based MCMC workflow. The R^2 guideline is appropriate for KNN or kriging proxies but not recommended for quadratic or cubic proxies since the polynomial proxies are inexact.

- The comparative study of CMOST and the proxy-based MCMC reveals that the efficiency, in term of solutions found per total simulations run, of CMOST is substantially higher than the proxy-based MCMC. However, in terms of accuracy, the former misses several history-matching solutions that are found by the latter.
- We also compare the direct MCMC with the proxy-based MCMC. It is found that the direct MCMC workflow is not recommended for history matching due to the enormous computational cost and it is preferable to use the proxy-based MCMC workflow.
- Finally, the AHM workflow and the EDFM preprocessor is successfully integrated using MATLAB and VBA. The combination the history matching performance of the former and the discrete fracture modelling power of the latter provides the integrated workflow that efficiently history matches a reservoir with complex fracture geometry.

6.2 RECOMMENDATIONS FOR FUTURE WORK

- The AHM workflow uses Metropolis Hasting (MH), a class of MCMC, as the sampling algorithm. Recently, there is a study claiming that Hamiltonian MCMC has better efficiency than MH. So it is recommend to compare the two MCMC algorithms.

- This study compares the performances of the four proxies: quadratic, cubic, KNN, and kriging. The comparative study should be extended to novel types of proxies such radial basis function, and deep learning.
- In this study, a proxy is built approximate the BHP RMSE. Alternatively, multiple proxies can be built to approximate BHP values for different time then BHP RMSE can be approximated by these proxy BHPs. Comparing the performance of the AHM workflows using one proxy (BHP RMSE) versus multiple proxies (BHP values) potentially leads to the improvement.
- If high performance cluster is available, the history matching of the field cases should be re-performed using kriging instead of KNN proxy.
- The complex-fracture case study using EDFM needs to disregard the permeability compliance because the current version of the EDFM preprocessor cannot handle the pressure dependent permeability. After the upgrade of the EDFM preprocessor to overcome this issue, the case study should be revisited and incorporated with the dynamic permeability.
- Although the study observes the impact of initial points on the performances of the AHM workflows. The result is still inconclusive due to the stochastic nature of the algorithms. It is recommended to redo the analysis multiple times and use the mean values as the basis to compare.
- The current version of the AHM workflow is written in MATLAB and VBA. For better portability and efficiency, the workflow is recommended to transfer to Python language in the future. This transfer will make the interaction between the AHM workflow and the EDFM preprocessor smoother since the latter is already written in Python.

- The AHM workflow is currently compatible with a commercial simulator (CMG) and the EDFM preprocessor. Integration of the AHM workflow with the in-house simulator (UTCOMP) that source code is accessible is recommended. Since the synergy will provide more flexibility to further develop the valuable and robust tool for reservoir characterization of tight reservoirs with complex fractures.

Glossary

NOMENCLATURE

AHM	=	Assisted history matching
AOP	=	Accuracy of proxy, %
BHP	=	Bottomhole pressure, psi
C_v	=	Coefficient of variation
DECE	=	Designed Exploration and Controlled Evaluation algorithm
DOE	=	Design of experiment
EnKF	=	Ensemble Kalman filter
KNN	=	K-nearest neighboring
LHD	=	Latin hypercube design
MCMC	=	Markov chain Monte Carlo
MH	=	Metropolis Hastings
PDF	=	Probability density functions
PPD	=	Posterior probability density (function)
RML	=	Randomized maximum likelihood
RMSE	=	Root-mean-square error
SRV	=	Stimulated rock volume
$A(\vec{\theta} \rightarrow \vec{\theta}^*)$	=	Transition probability to move from $\vec{\theta}$ to $\vec{\theta}^*$ vectors in MCMC
B	=	Formation volume factor, bbl/STB
c_{ij}	=	Covariance for kriging
C_t	=	Total compressibility, 1/psi
$\vec{d}$	=	Time-series of actual response, or observed data
d	=	Actual response, or observed data
$\vec{h}$	=	Displacement between two uncertain-parameter vectors
k	=	Permeability, md
k_{row}	=	Relative permeability to oil in oil-water system
k_{rw}	=	Relative permeability to water in oil-water system
p	=	Reservoir pressure, psi
$P(\vec{\theta} \mid \vec{z})$	=	Likelihood function of a random variable $\vec{\theta}$ to occur given observed data $\vec{z}$
q	=	Surface oil rate, STB/day
R^2	=	Determination index
t	=	Time, day
x_f	=	Fracture half-length, ft
$\vec{z}$	=	Time-series of actual response, or observed data
y	=	Horizontal distance perpendicular to a vertical fracture, ft
z	=	Actual response, or observed data
$\hat{z}$	=	Proxy response

β	=	Regression coefficient for a polynomial
ϕ	=	Porosity
γ	=	Permeability compliance, psi^{-1}
$\hat{\gamma}$	=	Variogram for kriging proxy
λ	=	Weight function, for KNN or kriging proxy
μ	=	Oil viscosity, cp
μ_0	=	Lagrange multiplier for kriging proxy
$\vec{\theta}$	=	Uncertain-parameter vector representing a reservoir-model realization
$\vec{\theta}^*$	=	Proposed uncertain-parameter vector in MCMC
σ^2	=	Variance in Gaussian distribution

SUBSCRIPT AND SUPERSCRIPT

actual	=	Value from actual reservoir simulation (as opposed to proxy)
D	=	Dimensionless
init	=	Initial
max	=	Maximum (upper boundary)
min	=	Minimum (lower boundary)
obs	=	Observed
proxy	=	Value from proxy model (as opposed to actual)
ref	=	Reference
sim	=	Simulated
*	=	Normalized

SI METRIC CONVERSION FACTORS

ft	×	3.048 e-01	=	m
ft^3	×	2.832 e-02	=	m^3
cp	×	1.0 e-03	=	Pa·s
psi	×	6.895 e+00	=	kPa
md	×	1e-15 e+00	=	m^2

References

Aanonsen, S.I., Naevdal G., Oliver, D.S., Reynolds, A.C., and Valles, B. 2009. The Ensemble Kalman Filter in Reservoir Engineering--a Review. *SPE Journal* **14** (3): 393 - 412. https://doi.org/10.2118/117274-PA

Barker, J.B., Cuypers, M., and Holden, L. 2001. Quantifying Uncertainty in Production Forecasts: Another Look at the PUNQ-S3 Problem. *SPE Journal* **6** (4): 433 - 441. SPE 74707. http://dx.doi.org/10.2118/74707-PA

Bello, R.O., Wattenbarger, R.A. 2008.Rate Transient Analysis in Naturally Fractured Shale Gas Reservoirs. CIPC/SPE Gas Technology Symposium Joint Conference, Calgary, Canada. 16-19 June. SPE 114591. https://doi.org/10.2118/114591-MS

Busby, D., Feraille, M. D., and Gervais-Couplet, V. 2009. Uncertainty Reduction By Production Data Assimilation Combining Gradual Deformation With Adaptive Response Surface Methodology. Society of Petroleum Engineers. SPE Europec/EAGE Annual Conference and Exhibition, Rome, 9 – 12 June. SPE 121274. http://dx.doi.org/10.2118/121274-MS

Behmanesh, H., Tabatabaie, S.H., Sureshjani, M.H., and Clarkson, C.R. 2014. Modification of the Transient Linear Flow Distance of Investigation Calculation for Use in Hydraulic Fracture Property Determination. SPE Unconventional Resources Conference, Woodlands, Texas, 1-3 April. SPE 168981. http://dx.doi.org/10.2118/168981-MS

Chen, C., Li, R., Gao, G., Vink, J.C., and Cao, R. 2016. EUR Assessment of Unconventional Asset Using Parallelized History Matching Workflow Together with RML Method. Unconventional Resources Technology Conference, San Antonio, Texas. 1-3 August. URTeC 2429986.

Chib S., and Greenberg E. 1955. Understanding the Metropolis-Hastings algorithm. *The American Statistician* **49** (4): 327-335. http://dx.doi.org/10.2307/2684568

Cipolla, C., Maxwell, S. and Mack, M. 2012. Engineering Guide to the Application of Microseismic Interpretation. SPE Hydraulic Fracturing Technology Conference, the Woodlands, Texas, 6 – 8 February. SPE 152165. https://doi.org/10.2118/152165-MS

Clarkson, C.R., Kovacs, J.W., Qanbari, F., Behmanesh, H., and Sureshjani, M.H. 2014. History-Matching and Forecasting Tight/Shale Gas Condensate Wells Using Combined Analytical, Semi-Analytical, and Empirical Methods. SPE/CSUR

Unconventional Resources Conference. Calgary, Canada, 30 September – 2 October. SPE 171593. https://doi-org.ezproxy.lib.utexas.edu/10.2118/171593-MS

CMG. 2014. CMOST User Guide. Computer Modeling Group Ltd.

Cover, T.M., and Hart, P. 1967. Nearest Neighbor Pattern Classification. *IEEE Transactions on Information Theory* **13** (1): 21-27. http://dx.doi.org/10.1109/TIT.1967.1053964

Elsakout, D, Christie, M., Lord, G. 2015. Multilevel Markov Chain Monte Carlo (MLMCMC) For Uncertainty Quantification. SPE North Africa Technical Conference and Exhibition, Cairo, Egypt, 14-16 September. SPE 175870. http://dx.doi.org/10.2118/175870-MS

Emerick, A.A., and Reynolds, A. C. 2010. EnKF-MCMC. SPE Europec/EAGE Annual Conference and Exhibition, Barcelona, Spain, 14 -17 June. SPE 131375. http://dx.doi.org/10.2118/131375-MS

Emerick, A.A., and Reynolds, A.C. 2011. Combining Sensitivity and Prior Information for Covariance Localization in the Ensemble Kalman Filter for Petroleum Reservoir Application. *Computational Geosciences*, **15**(2): 251 – 269

Evensen, G., Hove, J., Meisingset, H.C., Reisso, E., Seim, K.S., Espelid, S.O. 2007. Using the EnKF for Assisted History Matching of a North Sea Reservoir Model. SPE Reservoir Simulation Symposium, Woodlands, Texas, 26 – 28 February. SPE 106184. http://dx.doi.org/10.2118/106184-MS

Everitt, B. 1998. The Cambridge Dictionary of Statistics. Cambridge, UK New York: Cambridge University Press.

Gao, G.,Vink, J.C., Chen, C., Tarrahi, M., and Khamra, M.T. 2016. Uncertainty Quantification for History Matching Problems with Multiple Best Matches Using a Distributed Gauss-Newton Method. SPE Annual Technical Conference and Exhibition, Dubai. 26-28 September. SPE 181611. http://dx.doi.org/10.2118/181611-MS

Gao, G., Zafari, M., and Reynolds, A. 2005. Quantifying Uncertainty for the PUNQ-S3 Problem in a Bayesian Setting with RML and EnKF. SPE Reservoir Simulation Symposium, the Woodlands, Texas, 31 January-2 February. SPE 93324. https://doi.org/10.2118/93324-MS

Goodwin, N. 2015. Bridging the Gap Between Deterministic and Probabilistic Uncertainty Quantification Using Advanced Proxy Based Methods. SPE Reservoir Simulation

Symposium, Houston, 23-25 February. SPE 173301. http://dx.doi.org.ezproxy.lib.utexas.edu/10.2118/173301-MS

Hajizadeh, Y., Christie, M.A., and Demyanov, V. 2010. History matching with differential evolution approach; a look at new search strategies. SPE EUROPEC/EAGE Annual Conference and Exhibition, Barcelona, Spain, 14-17 June. SPE 130253. http://dx.doi.org/10.2118/130253-MS

Hamdi, H., Sousa, M.,C., Behmanes, H, and Thompson, A. 2017. Bayesian History-Matching and Probabilistic Forecasting for Tight and Shale Wells. SPE Unconventional Resources Conference, Calgary, Canada, 15-16 February. SPE 185082. http://dx.doi.org/10.2118/185082-MS

Isaaks, E.H., and Srivastava, R.M. 1989. An Introduction to Applied Geostatistics, first edition, Chap. 12, 287-332. New York: Oxford University Press.

Johnston K., Ver Hoef, J.M., and Krivoruchko, L., N. 2002a. The principles of geostatistical analysis. In ArcGIS 9: Using ArcGIS Geostatistical Analyst, Chap. 3, 53-74. Redlands, CA: ESRI.

Kalman, R.E. 1960. A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME – Journal of Basic Engineering* **82** (Series D): 35 – 45.

Landa, J.L., and Guyaguler, B. 2003. A Methodology for History Matching and the Assessment of Uncertainties Associated with Flow Prediction. SPE Annual Technical Conference and Exhibition, Denver. 5-8 October. SPE 84465. http://dx.doi.org/10.2118/84465-MS

Li, G., and Reynolds, A.C. 2009. Iterative Ensemble Kalman Filters for Data Assimilation. *SPE Journal* **14** (3): 496 – 505. SPE 109808. http://dx.doi.org/10.2118/109808-PA

Liu, N., and Oliver, D.S. 2003. Evaluation of Monte Carlo Methods for Assessing Uncertainty. SPE Journal 8(2): 188 – 195. SPE 84936. https://doi-org.ezproxy.lib.utexas.edu/10.2118/84936-PA

Ma, X., Al-Harbi, M., Datta-Gupta, A., and Efendiev, Y. 2006. A Multistage Sampling Method for Rapid Quantification of Uncertainty in History Matching Geological Models. SPE Annual Teleconference and Exhibition, San Antonio, Texas, 24 - 27 September. SPE 102476. http://dx.doi.org/10.2118/102476-MS

Shakiba, M., and Sepehrnoori, Kami. 2015. Using Embedded Discreted Fracture Model (EDFM) and Microseismic Monitoring Data to Charterize the Complex Hydraulic

Fracture Networks. SPE Annual Technical Conference and Exhibition, Houston, , 28-30 September. SPE 175142. https://doi.org/10.2118/175142-MS

Matheron, G. 1962. Traité de Géostatistique appliquée, Tome I. Mémoires du Bureau de recherches géologiques et minières., No. 14, Editions Technip. Paris.

Mohamed, L., Christie, M. A., and Demyanov, V. 2009. Comparison of Stochastic Sampling Algorithms for Uncertainty Quantification. Society of Petroleum Engineers. SPE Reservoir Simulation Symposium, the Woodlands, Texas, 2 - 4 February. SPE 119139. http://dx.doi.org/10.2118/119139-MS

Nejadi, S., Leung, J.Y., Trivedi, J.J., and Virues, C. 2015. Integrated Characterization of Hydraulically Fractured Shale-Gas Reservoirs – Production History Matching. SPE Res Eval & Eng 18 (4): 70 - 82 SPE 171664. http://dx.doi.org/10.2118/171664-PA

Olalotiti-Lawal, F., and Datta-Gupta, A. 2015. A Multi-Objective Markov Chain Monte Carlo Approach for History Matching and Uncertainty Quantification. SPE Annual Technical Conference and Exhibition, Houston 28 – 30 September. http://dx.doi.org/10.2118/175144-MS

Oliver, D.S., Reynolds, A.C., and Liu, N. 2008. Inverse Theory for Petroleum Reservoir Characterization and History Matching. New York: Cambridge University Press.

Samandarli, O., Al Ahmadi, H.A., Watternbarger, R.A., A Semi-Analytic Method for History Matching Fractured Shale Gas Reservoirs. 2011. SPE Western North American Region Meeting, Anchorage, Alaska, USA 7-11 May. SPE 144583. https://doi-org.ezproxy.lib.utexas.edu/10.2118/144583-MS

Schaaf, T., Coureaud, B., Labaf, N., and Busby, D. 2009. Using Experimental Designs, Assisted History-Matching Tools, and Bayesian Framework to Get Probabilistic Gas-Storage Pressure Forecasts. SPE Europec/EAGE Annual Conference and Exhibition, Rome, 9-12 June. SPE113498. https://doi.org/10.2118/142896-MS

Slotte, P.A., and Smorgrav, E. 2008. Response Surface Methodology Approach for History Matching and Uncertainty Assessment of Reservoir Simulation Models. SPE Eurospce/EAGE Annual Conference and Exhibition, Rome, 9-12 June. SPE 113390. http://dx.doi.org/10.2118/113390-MS

Taleghani, A.D., and Olson J.E. 2014. How Natural Fractures Could Affect Hydraulic-Fracture Geometry. *SPE Journal* **19** (1): 161 – 171. SPE 167608. https://doi.org/10.2118/167608-PA

Tavakoli, R., Srinivasan, S., and Wheeler, M.F. 2014. Rapid Updating of Stochastic Models by Use of an Ensemble-Filter Approach. *SPE Journal* **19** (3): 500 – 513. SPE-163673. https://doi.org/10.2118/163673-PA

Tierney L. 1994. Markov chains for exploring posterior distributions. *Annual of Statistics* **22** (4): 1701-1762. http://dx.doi.org/10.1214/aos/1176325750

Vazquez, O., Young, C., Demyanov, V., and Arnold, D. 2015. Produced-Water-Chemistry History Matching in the Janice Field. *SPE Journal* **18** (4): 564 - 576. SPE 164903. http://dx.doi.org/10.2118/164903-PA.

Wantawin, M. 2016. A Probabilistic Workflow for Uncertainty Analysis Using a Proxy-Based Approach Applied to Tight Reservoir Simulation Studies. MS thesis. The University of Texas at Austin, Austin, Texas (August 2016). www.pge.utexas.edu/images/pdfs/theses16/wantawin.pdf.

Warpinski, N.R., Mayerhofe, M.J., Agarwal, K., and Du, J. 2013. Hydraulic-Fracture Geomechanics and Microseismic-Source Mechanisms. *SPE Journal* **18** (4): 766 – 780. SPE 158935. https://doi.org/10.2118/158935-MS

Warpinski, N.R., Mayerhofer, M.J., Vincent, M.C., Cipolla, C.L., and Lolon, E.P. 2009. Stimulating Unconventional Reservoirs: Maximizing Network Growth While Optimizing Fracture Conductivity. *JCPT* **48** (10): 39 – 51. SPE 114173. https://doi.org/10.2118/114173-PA

Xu, Y., Cavalcante Filho, J.S.A.,Yu., W., and Sepehrnoori, K. 2017. Discrete-Fracture Modeling of Complex Hydraulic-Fracture Geometries in Reservoir Simulators. *SPE Journal* **20** (2): 403 – 422. SPE 183647. https://doi-org.ezproxy.lib.utexas.edu/10.2118/183647-PA

Yang, C., Nghiem, L., Erdle, J., Moinfar, A., Fedutenko, E., Li, H., Mirzabozorg, A., and Card, C. 2015. An Efficient and Practical Workflow for Probabilistic Forecasting of Brown Fields Constrained by Historical Data. SPE Annual Technical Conference and Exhibition, Houston, 28 - 30 September. SPE 175122. http://dx.doi.org/10.2118/175122-MS

Yeten, B., Castellini, A., Guyaguler, B., and Chen, W. H. 2005. A Comparison Study on Experimental Design and Response Surface Methodologies. Society of Petroleum Engineers. SPE Reservoir Simulation Symposium, Texas, 31 January - 2 February. SPE 93347. http://dx.doi.org/10.2118/93347-MS

Yin, J., Xie, J., Datta-Gupta A., and Hill, A.D. 2011. Improved Characterization and Performance Assessment of Shale Gas Wells by Integrating Stimulated Reservoir

Volume and Production Data. SPE Eastern Regional Meeting, Columbus, Ohio, USA, 17 - 19 August. SPE 148969. https://doi-org.ezproxy.lib.utexas.edu/10.2118/148969-MS

Zhang, Z., and Fassihi, M.R. 2013. Uncertainty Analysis and Assisted History Matching Workflow in Shale Oil Reservoirs. Unconventional Resources Technology Conference, Denver, 12-18 August. SPE 168795. http://dx.doi.org/10.1190/URTEC2013-002

Zubarev, D.I. 2009. Pros and Cons of Applying Proxy-Models as a Substitute for Full Reservoir Simulations. SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana, 4-7 October. SPE 124815. http://dx.doi.org/10.2118/124815-MS